37181- 1927.

1909–1910

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

EDITED BY

ORMOND STONE W. E. BYERLY F. S. WOODS
E. V. HUNTINGTON J. K. WHITTEMORE
ELIJAH SWIFT

PUBLISHED UNDER THE AUSPICES OF HARVARD UNIVERSITY

SECOND SERIES Vol. 11

FOR SALE BY
THE PUBLICATION OFFICE OF HARVARD UNIVERSITY
2 University Hall, Cambridge, Mass., U. S. A

London: LONGMANS, GREEN & Co.
39 Paternoster Row

Leipzig: OTTO HARRASSOWITZ
Querstrasse 14

THEORY OF FLOATING TUBES

By Frank Gilman

Article 1. It is proposed to treat of the theory of floating tubes or rods as applied to the measurement of the velocity of water in open channels; but, before doing so, we shall give a brief historical sketch of the use of this method in water measurements.

The first public mention of the method, so far as known, was made by T. A. Mann, a member of the Imperial and Royal Academy of Sciences at Brussels. He communicated a paper to the President of the Royal Society of London, Joseph Banks, who read it before the Society on June 24, 1779. The subject of the paper was "The Hydraulics of Rivers and Canals," and in it a method is described for measuring the velocity of water in canals by means of wooden rods, or poles, of a length somewhat less than the depth of the water, while at the lower end are suspended as many small weights as may be necessary to keep the rod in a vertical position. He advises that a small straight wire be fastened to the center of that end of the rod which projects from the water so that it may indicate the deviations of the rod from a vertical position, and enable inferences to be made in regard to the relative velocities of the water at different depths.

Mr. Mann speaks of this method as the best and simplest of which he knows for measuring the velocity of the water in canals and rivers. He refers to it as if it were a well known method, but gives no example of its application. It is probable, however, that he had used it, for he says that he had long lived in a country that abounded in canals, and had been much employed in matters relating to hydraulics.

The results of the application of this method were first published by C. R. T. Krayenhoff in Amsterdam, Holland, in 1813. His work gives an account of his observations on the hydrography and topography of Holland. The floats that he used were wooden poles loaded with lead at the bottom, and carrying copper floats at the surface. The next application of the method was made by M. de Buffon, who in 1821 gauged the Tiber by the use of bundles of rods loaded at the lower ends and extending from the surface nearly to the bottom.

In 1835 Destrem gauged the Neva by the same method.

In measuring the discharge of small canals Hirn used light covered frames, so arranged as to be at right angles to the current, and of such extent that they nearly filled the whole cross-section of the stream, and consequently gave, at one reading, the approximate mean velocity at that section.

In 1852 James B. Francis, of Lowell, Mass., made most elaborate experiments on measuring the discharge of canals by the use of loaded tubes. These experiments were supplemented by others made in 1856, and a full account of all of them can be found in "The Lowell Hydraulic Experiments," a second edition of which was published by Van Nostrand in 1868. Mr. Francis compared the discharge given by the use of tubes with the same quantity passing over a weir and determined by weir measurements, and found that the difference was generally less than two per cent.

The tubes used were hollow tin cylinders, two inches in diameter, soldered together, with a solid cylindrical piece of lead, of the same diameter, at the lower end. The tubes ranged in length from 6 to 10 feet, and the centers of gravity of the longer tubes were about 1.9 feet from the lower ends. Two beams were laid across the canal, at right angles to the current and 70 feet apart. These constituted the upper and lower transit stations. The time occupied by the tube in passing from one station to the other, was determined by means of a stop watch, or chronometer; and from this and the known distance the velocity of the tube was calculated. The up-stream side of each beam was figured from one end to the other, so that the distance from the left shore (looking down stream) at which the tube passed each beam could be noted. The results were plotted on cross-section paper, calling the mean distance from the left shore the abscissa, and the velocity the ordinate. Between the points thus obtained a curve was drawn so that the sum of the vertical distances from the curve of the plotted points which are above, should be equal to the sum of the vertical distances of the points which are below the curve, or so that the sum of the positive errors should be equal to the sum of the negative errors.

From the mean velocity curve, thus obtained, readings were taken at intervals corresponding to one foot each in the width of the stream. The sum of these readings multiplied by the mean depth gave the discharge.

Very elaborate experiments in measuring the discharge of streams by the use of rods were made by Capt. Allan Cunningham near Roorkee, India, on the Ganges Canal. These experiments were begun in 1874 and continued, with some intermissions, until 1879. Capt. Cunningham found that rods moved

more steadily than any other sort of float, that they gave the result more rapidly, were more easily handled and less delicate, being simple in construction. He recommends that for measurements of mean velocity past a vertical, the rods should supersede all other instruments in cases favorable to their use. The conditions favorable to their use are that the cross-section and declivity of the stream should be uniform for a considerable distance, that the bottom should be free from obstructions, and that the depth should not exceed 15 feet.

M. A. Graëff, in his "Traité d'hydraulique," published in 1883, expresses a similar opinion in regard to the merits of the loaded tube, or rod. He states that Italian engineers had adopted the method of measuring the mean velocity by means of rods, and that he himself had used it in gauging the discharge of the Loire and its tributaries.

The Mississippi River experiments of Messrs. Humphreys and Abbot remain to be mentioned, which are the most important of all, as far as this paper is concerned, since by means of them the truth of our fundamental formula, $v = a + bx + cx^2$, was first demonstrated, in which v denotes the velocity of the water at the depth x, the total depth being unity, while a, b, and c are constants determined by experiment. The experiments were made at Carrolton and Baton-Rouge, Louisiana, in 1851. The mean depth of the river was about 82 feet. This depth was divided into 10 equal parts, and at each proportional depth 222 observations of the velocity were made, and their mean taken as the true velocity.

The results are tabulated on page 244 of the "Report of the Mississippi River," published at Washington in 1876. The mean results are given further on in this article, where it is seen that they satisfy the parabola equation with considerable exactness. Long tubes, of course, could not be used in measuring the velocity at different depths, nor in gauging the discharge for a river of such depth as the Mississippi. The apparatus used was a double float, consisting of a surface-float, a sub-float, and a connecting cord. The surface-float was of cork, 5 inches long, 1 inch thick, and submerged to a depth of $1\frac{1}{2}$ inches. Its weight, therefore, was not more than one-fourth of a pound.

A wire one foot in length, and carrying a small flag, was inserted in this float. The sub-float was a keg, open at both ends, beveled at the lower edge, and weighted with strips of lead to keep it in a vertical position. The weight of the keg, with the ballast, was about 9 pounds. Its diameter was 10 inches and its height 15 inches, thickness of staves three-eighths of an inch. The connecting cord was of hemp and one-tenth of an inch in diameter.

Its weight when stretched to its full length of 90 feet, was one-half a pound. These experiments have been severely criticized, on the ground that the sub-velocity measurements, especially those at great depths, did not truly represent the velocities at these points, on account of the disturbing influences due to the surface float and connecting cord. But these criticisms do not seem to be well founded, as a little calculation will show that the vis viva of the surface-float and cord due to the difference between the velocity of the water at the bottom, and the mean velocity of the water surrounding the cord, would be less than one-tenth of one per cent. of the total vis viva of the keg. Another result of the Mississippi River experiments was the conclusion that there is a nearly constant ratio between the mid-depth velocity and the mean velocity past a vertical line from the surface to the bottom. This follows however, from the form of the equation which gives the relation between the velocities at different depths, viz., $v = a + bx + cx^2$. Calling v_m the mean velocity past a vertical line, and $v_{\frac{1}{2}}$ the mid-depth velocity, we evidently have

$$\frac{v_m}{v_{\frac{1}{2}}} = \frac{a + \frac{b}{2} + \frac{c}{3}}{a + \frac{b}{2} + \frac{c}{4}};$$

substituting the values of a, b, and c as deduced from the Mississippi River experiments, viz:

$$a = +\,3.1952, \qquad b = +\,0.4424, \qquad c = -\,0.7652,$$

we find $v_m/v_{\frac{1}{2}} = 0.980$.

If this ratio, 0.98, be applied to Capt. Cunningham's first 46 series of experiments, and each of the mid-depth velocities be multiplied by 0.98, the results will differ from the v_m of his experiments by less than one per cent in the majority of cases. The following table gives a synopsis of the results of the Mississippi River experiments:

Relative Depth	Observed velocity in feet per sec.	Velocity by formula
x	v	$v = a + bx + cx^2$.
0	3.1950	3.1952
.1	3.2299	3.2318
.2	3.2532	3.2531
.3	3.2611	3.2591
.4	3.2516	3.2497
.5	3.2282	3.2251
.6	3.1807	3.1852
.7	3.1266	3.1299
.8	3.0594	3.0594
.9	2.9759	2.9735

ARTICLE 2. In discussing the theory of floating tubes different results will be obtained according to the law assumed to hold true for the resistance of fluids. Some engineers assume that the resistance of fluids varies as the first power of the velocity, and consequently they take it for granted that the velocity of a floating tube is the same as the mean velocity of the water in the same vertical line of the same length as the tube. It would follow from this that the proper length of a tube, theoretically, for measuring the velocity of a current, is the same as the depth. But since in practice the length of the tube must be less than the depth, it is inferred that the velocity of the tube is always greater than the mean velocity of the current, in consequence of the slower motion of the water along the bottom, and tables of corrections are prepared to be applied to the observed velocity of the tube, which corrections are always negative. It is easily shown by analysis that this assumption and practice would be correct if the impulse and resistance of fluids varied directly as the velocity. But according to the law accepted by the great majority of experimenters, the impulse and resistance vary as the square of the velocity; and not only has this been demonstrated by experiment, but it has been shown to be true from theoretical considerations.*

* For a theoretical proof, see Weisbach's *Mechanics*, Vol. 1, Article 498.

We will now investigate the conditions for determining the velocity of a tube when immersed vertically in a stream of water, on the hypothesis that the resistance generated by the motion of a solid body in a still fluid is proportional to the square of the velocity, and that when the fluid itself has motion the resistance is proportional to the square of the relative velocity of the solid and fluid.

In the case of a vertical tube borne along by the current some parts of the tube will move faster than the adjacent fluid, and some slower. The parts that move faster will meet a force of resistance, and those that move slower a force of acceleration, each of which will be proportional to the square of the relative velocity of that portion of the tube and adjacent fluid.

We shall discuss in this article the case in which the velocity of the tube is less than the velocity of the water at the surface, as represented in figures 1 and 2, in which LP denotes the tube, AB the surface of the water, ACI the parabolic curve of velocity, EF the velocity v, corresponding to the depth BF as x, the relation of v and x being expressed by the formula, $v = a + bx + cx^2$.

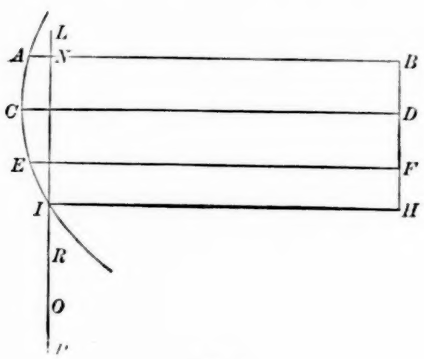


FIG. 1.

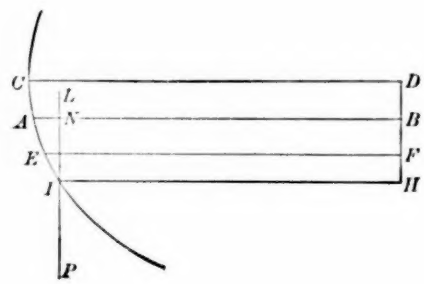


FIG. 2.

Figure 1 represents the case in which the axis of the parabola, CD, (which is the maximum velocity line) is below the surface of the water, while figure 2 represents the case when this axis is above the surface.

In the former case b/c is negative, in the latter positive. When the tube has attained a state of uniform motion, the sum of the forces of resistance will be equal and opposite to the sum of the forces of acceleration, a condition expressed by the following equation :

$$\int_0^h (v - v')^2 dx = \int_h^l (v' - v)^2 dx,$$

in which h denotes the distance from the surface to the point where the velocity of the water is equal to the velocity of the tube, and is represented in the figures by BH; v' denotes the velocity of the tube, and is equal to $a + bh + ch^2$, while l is the length of the submerged portion of the tube, taken on the same scale as x, the total depth being unity.

Substituting the values of v and v', as above given, performing the indicated operations, and arranging the results with reference to the powers of h, we have

$$(1) \quad \frac{16}{15} h^5 + \left(\frac{5}{3}\frac{b}{c} - l\right)h^4 + \left(\frac{2}{3}\frac{b^2}{c^2} - 2\frac{b}{c}l\right)h^3 + \left(\frac{b}{c}l^2 + \frac{2}{3}l^3 - \frac{b^2}{c^2}l\right)h^2$$
$$+ \left(\frac{b^2}{c^2}l^2 + \frac{2}{3}\frac{b}{c}l^3\right)h = \frac{1}{3}\frac{b^2}{c^2}l^3 + \frac{b}{2c}l^4 + \frac{l^5}{5}.$$

Let $h = rl$, and substitute this value of h in the above equation, then, dividing through by l^5, and representing b/cl by s, we obtain the following for the final equation:

$$(2) \quad \frac{16}{15} r^5 + \left(\frac{5}{3}s - 1\right)r^4 + \left(\frac{2}{3}s^2 - 2s\right)r^3 + \left(s - s^2 + \frac{2}{3}\right)r^2$$
$$+ \left(\frac{2}{3}s + s^2\right)r - \frac{s^2}{3} - \frac{s}{2} - \frac{1}{5} = 0.$$

It is easily shown, by Sturm's Theorem, that this equation will give only one real value for r corresponding to any real value of s; for instance let $s = -1$, then the equation becomes

$$r^5 - 2.5r^4 + 2.5r^3 + 0.625r^2 + 0.3125r - 0.03125 = 0.$$

Call this equation X, and its first derived polynomial X_1. Then finding the greatest common divisor of X and X_1, denoting the successive remainders, with their signs changed, by R, R_1, R_2, etc., and writing the results in two rows, we have

X	X_1	R	R_1	R_2
r^5	r^4	$-r^2$	$-r$	$+1.55164$

Each term in the second row denotes the first term of the equation designated by the symbol immediately above.

Substituting successively $-\infty$ and $+\infty$ for r in the equations X, X_1, R, R_1, and R_2, we have the following signs:

for $r = -\infty$	$- + - + +$	three variations;
for $r = +\infty$	$+ + - - +$	two variations;

the equation has therefore one real root.

Next substitute $+1$ for s in equation (2), and by a similar process we obtain the following results:

	X	X_1	R	R_1	R_2	R_3	
	r^5	r^4	r^3	$-r^2$	r	$+0.58$	
for $r = -\infty$	$-$	$+$	$-$	$-$	$-$	$+$	three variations;
for $r = +\infty$	$+$	$+$	$+$	$-$	$+$	$+$	two variations.

Again the equation has one real root.

The following tables give the values of r corresponding to different values of s, negative values of s being given by the first table, and positive values by the second.

TABLE 1 s negative		TABLE 2 s positive					
s	r	s	r	s	r	s	r
—		+		+		+	
0	0.6105	0					
.1	.6200	.1	0.6023	1.1	0.5577	2.1	0.5397
.2	.6306	.2	.5952	1.2	.5552	2.2	.5385
.3	.6431	.3	.5890	1.3	.5529	2.3	.5374
.4	.6577	.4	.5834	1.4	.5508	2.4	.5363
.5	.6744	.5	.5785	1.5	.5489	2.5	.5353
.6	.6936	.6	.5741	1.6	.5471	2.6	.5343
.7	.7150	.7	.5701	1.7	.5454	2.7	.5334
.8	.7364	.8	.5665	1.8	.5438	2.8	.5326
.9	.7523	.9	.5633	1.9	.5424	2.9	.5318
1.0	.5000	1.0	.5604	2.0	.5410	3.0	.5310

The following example will illustrate the application of these tables: Let the relative length, l, of a tube be 0.925; it is required to find its velocity when the values of a, b, and c are the same as those found in the Mississippi River experiments, viz:

$$a = +3.1950, \qquad b = +0.4424, \qquad c = -0.7652.$$

We have

$$s = \frac{b}{cl} = -\frac{0.4424}{0.7652 \times 0.925} = -0.625.$$

With the argument $s = -0.625$, we find from table 1, by interpolation, $r = 0.6989$, whence

$$h = rl = 0.6989 \times 0.925 = 0.6465;$$

and the velocity of the tube,

$$v' = a + bh + ch^2 = 3.161.$$

Assuming that the impulse and resistance of fluids varies as the first power of the velocity, the problem would be solved as follows: Since in this case the velocity of the tube would be the same as the mean velocity of the current taken in a vertical line from the surface to the depth 0.925, we should have the following expression for the velocity of the tube:

$$v' = \frac{1}{0.925}\int_0^{0.925}(a + bx + cx^2)\,dx = 3.181.$$

ARTICLE 3. Equations 1 and 2 and the preceding tables apply when the

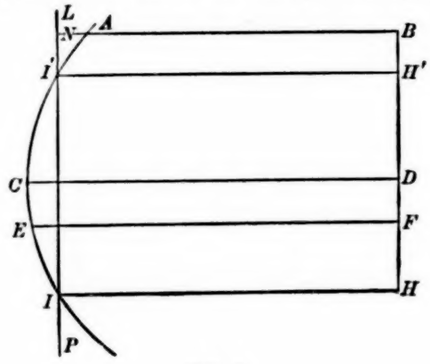


FIG. 3.

surface velocity of the water is greater than the tube velocity; but when the surface velocity is less than the tube velocity, a different equation is required

to express the conditions of equilibrium. Figure 3 gives a graphical representation of this case, in which the velocities are represented, as before, by horizontal lines drawn from BH to the parabolic curve. The surface velocity is also represented by AB, and the tube velocity by each of the lines IH and $I'H'$. EF denotes the velocity, v, of the water at any depth $BF = x$. As the axis of the tube, LP, intersects the velocity curve in the two points I and I', it is evident that the velocity of the water at each of these points is the same as the velocity of the tube. Putting $NI = h$ and $NI' = h_1$, we have the following two expressions for the velocity of the tube:

$$v' = a + bh + ch^2 = a + bh_1 + ch_1^2.$$

Since from N to I' and from I to P the velocity of the water is less than the velocity of the tube, while from I' to I the velocity of the water is greater than the tube velocity, we have, according to the principles explained in Article 3, the following equation of condition:

$$\int_0^{h_1}(v' - v)^2 dx + \int_h^l (v' - v)^2 dx = \int_{h_1}^h (v - v')^2 dx\,;$$

substituting for v and v' their values, integrating and arranging the results with reference to the powers of h, and remembering that $h + h_1 = -b/c$,* we have

$$\begin{aligned}(3)\quad & \frac{32}{15}h^5 + \left(\frac{32}{6}\frac{b}{c} - l\right)h^4 + \left(\frac{16}{3}\frac{b^2}{c^2} - 2\frac{b}{c}l\right)h^3 + \left(\frac{8}{3}\frac{b^3}{c^3} - \frac{b^2}{c^2}l + \frac{b}{c}l^2 + \frac{2}{3}l^3\right)h^2 \\ & + \left(\frac{2}{3}\frac{b^4}{c^4} + \frac{b^2}{c^2}l^2 + \frac{2}{3}\frac{b}{c}l^3\right)h + \frac{b^5}{15c^5} - \frac{b^2}{c^2}\frac{l^3}{3} - \frac{b}{c}\frac{l^4}{2} - \frac{l^5}{5} = 0.\end{aligned}$$

Writing s for b/cl, and r for h/l, and dividing through by l^5, we obtain

$$\begin{aligned}(4)\quad & \frac{32}{15}r^5 + \left(\frac{32}{6}s - 1\right)r^4 + \left(\frac{16}{3}s^2 - 2s\right)r^3 + \left(\frac{8}{3}s^3 - s^2 + s + \frac{2}{3}\right)r^2 \\ & + \left(\frac{2}{3}s^4 + s^2 + \frac{2}{3}s\right)r + \frac{s^5}{15} - \frac{s^2}{3} - \frac{s}{2} - \frac{1}{5} = 0.\end{aligned}$$

---

* In order to prove this, we find, by differentiating the equation $v = a + bx + cx^2$, that the depth of maximum velocity is $x_1 = -\frac{b}{2c} = BD$ in figure 3, or $2\,BD = -\frac{b}{c} = BH + BH'$ $= h + h_1$.

This equation will give only one real value of r corresponding to any real value of s; proceeding as before we find by Sturm's Theorem when $s = -0.75$,

	X	X_1	R	R_1	R_2	R_3	
	r^5	r^4	r^3	$-r^2$	r	$+1.36$	
for $r = -\infty$	$-$	$+$	$-$	$-$	$-$	$+$	three variations;
for $r = +\infty$	$+$	$+$	$+$	$-$	$+$	$+$	two variations.

The equation has one real root.

The following table gives the values of r corresponding to different values of s:

TABLE 3 s negative

$-s$	r	$-s$	r	$-s$	r	$-s$	r	$-s$	r	$-s$	r
0.70	0.7149	0.80	0.7372	0.90	0.7645	1.00	0.8053	1.10	0.8644	1.20	0.9372
.71	.7171	.81	.7395	.91	.7679	1.01	.8104	1.11	.8713	1.21	.9449
.72	.7193	.82	.7419	.92	.7714	1.02	.8157	1.12	.8783	1.22	.9526
.73	.7215	.83	.7444	.93	.7750	1.03	.8212	1.13	.8854	1.23	.9603
.74	.7237	.84	.7470	.94	.7787	1.04	.8268	1.14	.8925	1.24	.9681
.75	.7259	.85	.7497	.95	.7826	1.05	.8326	1.15	.8997	1.25	.9759
.76	.7281	.86	.7524	.96	.7867	1.06	.8385	1.16	.9070	1.26	.9837
.77	.7303	.87	.7552	.97	.7910	1.07	.8446	1.17	.9144	1.27	.9915
.78	.7326	.88	.7582	.98	.7955	1.08	.8510	1.18	.9219	1.28	.9993
.79	.7349	.89	.7613	.99	.8003	1.09	.8576	1.19	.9295	1.29	1.0071

The velocity of a tube of given length, when $-b/c$ is greater than $2/3$, can be found by the use of table 3, in the same manner as already shown in connection with tables 1 and 2.

ARTICLE 4. The problem of finding the velocity of a tube of given length is of less practical importance than that of finding the length of tube whose velocity shall be the same as the true mean velocity of the current in the same vertical line, and taken from the surface to the bottom of the stream.

We will call this the equivalent length, and designate it by L. Resuming equation (1), and arranging the terms with reference to the powers of l, and writing L for l, we have

$$(5)\quad \frac{L^5}{5} + \frac{b}{2c}L^4 + \left(\frac{b^2}{3c^2} - \frac{2}{3}\frac{e}{c}\right)L^3 - \frac{e}{c}\frac{b}{c}L^2 + \frac{e^2}{c^2}L - \frac{16}{15}h^5 - \frac{5}{3}\frac{b}{c}h^4 - \frac{2}{3}\frac{b^2}{c^2}h^3 = 0,$$

in which $e/c = (b/c)\,h + h^2$. In order to show that this equation has but one real root, substitute in it $-1/2$ for b/c, and we have

	X	X_1	R	R_1	R_2	R_3	
	L^5	L^4	L^3	$-L^2$	$-L$	$+0.06$	
for $L = -\infty$	$-$	$+$	$-$	$-$	$+$	$+$	three variations;
for $L = +\infty$	$+$	$+$	$+$	$-$	$-$	$+$	two variations.

The equation has one real root. In order that L in this equation may represent the equivalent length, h must be determined by the condition that the velocity of the water at the depth h shall be the same as the mean velocity of the current, taken in the same vertical line as the axis of the tube, and from the surface to the bottom of the stream. But this mean velocity is given by the following expression:

$$\int_0^1 v\,dx = \int_0^1 (a + bx + cx^2)\,dx = a + \frac{b}{2} + \frac{c}{3}.$$

Therefore the condition for determining h is

$$a + bh + ch^2 = a + \frac{b}{2} + \frac{c}{3}$$

or

$$h = -\frac{b}{2c} + \sqrt{\left(\frac{b}{2c}\right)^2 + \frac{b}{2c} + \frac{1}{3}}.$$

Using this value of h in equation (5) we obtain the following values of the roots, or equivalent lengths, corresponding to different values of b/c, and corresponding to the case discussed in Article 3.

TABLE 4

$\frac{b}{c}$	L	$\frac{b}{c}$	L	$\frac{b}{c}$	L	$\frac{b}{c}$	L
−		+		+		+	
0	0.946			1.0	0.967	2.0	0.976
.1	.942	0.1	0.949	1.1	0.968	2.1	0.977
.2	.939	.2	.951	1.2	.969	2.2	.977
.3	.935	.3	.954	1.3	.970	2.3	.978
.4	.930	.4	.956	1.4	.971	2.4	.978
.5	.927	.5	.958	1.5	.972	2.5	.979
.6	.925	.6	.960	1.6	.973	2.6	.979
		.7	.962	1.7	.974	2.7	.980
		.8	.964	1.8	.975	2.8	.980
		.9	.966	1.9	.976	2.9	.981

ARTICLE 5. We will next determine L when the velocity of the tube is greater than the surface velocity, which is the case discussed in article 3. The condition in this case, expressed analytically, is as follows:

$$a + bh + ch^2 > a, \qquad \text{or} \qquad -\frac{b}{c} > h,$$

and from the formula for h as given in article 4, we find that when $b/c = -2/3$, $h = 2/3$, and that when b/c is negative and numerically greater than $2/3$, $-b/c > h$. Therefore it is evident that for such values of b/c equation (3) must be used to determine L.

Arranging this equation with reference to the powers of l, and writing L for l, we have

$$\begin{aligned} &\frac{L^5}{5} + \frac{b}{2c}L^4 + \left(\frac{b^2}{3c^2} - \frac{2}{3}\frac{e}{c}\right)L^3 - \frac{e}{c}\frac{b}{c}L^2 + \frac{e^2}{c^2}L \\ (6)\quad &-\frac{2e^2}{c^2}\left(h - h_1\right) + \frac{2e}{c}\frac{b}{c}\left(h^2 - h_1^2\right) + \left(\frac{4e}{3c} - \frac{2}{3}\frac{b^2}{c^2}\right)\left(h^3 - h_1^3\right) \\ &-\frac{b}{c}\left(h^4 - h_1^4\right) - \frac{2}{5}\left(h^5 - h_1^5\right) = 0, \end{aligned}$$

in which $e/c = bh/c + h^2$, and $h_1 = -b/c - h$.

For any value of L given by equation (6), and corresponding to a given value of b/c and the following values of h and h_1, viz:

$$h = -\frac{b}{2c} + \sqrt{\left(\frac{b}{2c}\right)^2 + \frac{b}{2c} + \frac{1}{3}},$$

$$h_1 = -\frac{b}{2c} - \sqrt{\left(\frac{b}{2c}\right)^2 + \frac{b}{2c} + \frac{1}{3}},$$

there will also be a value of L given by equation (5), in which h_1 is to be used instead of h. It is seen by inspection that the coefficients of the corresponding powers of L in equations (5) and (6) are identical, for a given value of b/c.

Moreover these coefficients will be the same whether h or h_1 be used; for we have $h + h_1 = -b/c$, and multiplying both members by $h - h_1$, and transposing, we find

$$\frac{b}{c}h + h^2 = \frac{b}{c}h_1 + h_1^2 = \frac{e}{c}.$$

It follows that for the same values of b/c, equations (5) and (6) will differ only in their absolute terms. When $b/c = -2/3$, one value of L is zero; that is a surface float will give the true mean velocity.

Applying Sturm's Theorem to equation (6), after substituting -0.8 for b/c, we have

	X	X_1	R	R_1	R_2	R_3	
	L^5	L^4	L^3	$L^2-L+0.45$			
for $L=-\infty$	$-$	$+$	$-$	$+$	$+$	$+$	three variations;
for $L=+\infty$	$+$	$+$	$+$	$+$	$-$	$+$	two variations.

The equation has one real root.

The following table gives the two values of L corresponding to the same value of b/c, the numbers in the second column being roots of equation (5), and those in the last column roots of equation (6):

TABLE 5

$\frac{-b}{c}$	L (5)	L (6)
0.7	0.050	0.929
.8	.196	.941
.9	.393	.951
1.0	.600	.929
1.01	.633	.918
1.02	.676	.900
1.03	.804	.804

ARTICLE 6. The method of determining the values of the constants, a, b, and c, in the formula $v = a + bx + cx^2$, will be briefly described. In the second column of the table at the end of article 1, are given the values of observed velocities at proportional depths, and from these data the equations of condition are written as follows:

$$
\begin{array}{lllllll}
a & & & & & = & 3.1950 \\
a & + & 0.1b & + & 0.01c & = & 3.2299 \\
a & + & .2b & + & .04c & = & 3.2532 \\
a & + & .3b & + & .09c & = & 3.2611 \\
a & + & .4b & + & .16c & = & 3.2516 \\
a & + & .5b & + & .25c & = & 3.2282 \\
a & + & .6b & + & .36c & = & 3.1807 \\
a & + & .7b & + & .49c & = & 3.1266 \\
a & + & .8b & + & .64c & = & 3.0594 \\
a & + & .9b & + & .81c & = & 2.9759
\end{array}
$$

The normal equations, formed from the above equations of condition, are as follows:

$$
\begin{array}{rlrlrll}
10a & + & 4.5b & + & 2.85c & = & 31.7616 \\
4.5a & + & 2.85b & + & 2.025c & = & 14.0896 \\
2.85a & + & 2.025b & + & 1.5332c & = & 8.8288
\end{array}
$$

The solution of these equations gives

$$a = +\,3.1952, \quad b = +\,0.4424, \quad c = -\,0.7652.$$

Article 7. It was stated in the first part of article 2 that the hypothesis of an impulse and resistance of fluids proportional to the first power of the velocity necessarily involves the assumption that the equivalent length of a tube is always equal to the total depth of the stream at the point of immersion.

As this case is very simple, we will give the analytical proof. The equation expressing the conditions of the equilibrium of the forces acting on the tube is

$$\int_0^h (v - v')dx = \int_h^l (v' - v)dx.$$

Substituting the values of v and v', integrating and reducing, we have

$$\frac{l^2}{3} + \frac{b}{2c}\, l = h^2 + \frac{b}{c}\, h.$$

In order that l, in this equation, may represent the equivalent length, we must have

$$h = -\frac{b}{2c} + \sqrt{\left(\frac{b}{2c}\right)^2 + \frac{b}{2c} + \frac{1}{3}}.$$

Substituting this value of h in the above equation, and writing L for l, we have

$$\frac{L^2}{3} + \frac{b}{2c}\, L = \frac{b}{2c} + \frac{1}{3},$$

whence $L = 1$. In the same manner it may be shown that the value of L derived from the equation,

$$\int_0^{h_1} (v' - v)dx + \int_h^l (v' - v)dx = \int_{h_1}^h (v - v')dx$$

is always equal to unity.

The value of the second root in this case is $L = -\left(1 + \frac{3}{2}\frac{b}{c}\right)$, and this is always positive and real and ranges from 0 to $\frac{1}{2}$.

When $b/c = -\,2/3$, its value is 0, which is the same value that obtains on the hypothesis of a resistance proportional to the square of the velocity.

Article 8. We have hitherto assumed that the tube in floating maintains a vertical position, and it is now necessary to prove that this is true within practical limits. For this purpose, we will develop a formula for finding the inclination of the tube. We shall consider only the case discussed in article 3, in which from the surface to the depth h the velocity of the water is greater than the velocity of the tube, while for depths exceeding h the velocity of the water is less than the tubular velocity.

Since in this case the moments of the forces acting on the tube all tend to produce rotation in the same direction, it will be the case in which the inclination of the tube is greatest. The case is represented in figure 1, the tube being denoted by LP; LN is the portion of the tube that projects out of the water, OP the weighted portion of the tube, and R its center of gravity; NP, the immersed portion of the tube, is the part that we have designated by l, or L; NI is h.

It is evident that from N to I the forces acting on the tube, at right angles to its axis, all act in the same direction. Call the sum of the moments of these forces taken with reference to I, M_1, and the sum of the moments of the forces acting in the opposite direction, from I to P, M_2. The sum of these two moments must be equal to the moment of stability of the tube, or $M_1 + M_2 = WC \sin f$, in which W is the weight of the tube, and is equal and opposite to the upward thrust of the displaced water, $C \sin f$ is the arm of the couple of these two forces, and f is the angle of deviation of the tube from the vertical. Whence

$$\sin f = \frac{M_1 + M_2}{WC}.$$

Prof. Rankine in his "Applied Mechanics," article 652, gives the following formula for the pressure of a current upon a solid body floating or immersed in it:

$$R = kD \frac{v^2}{2g} A,$$

in which R is the pressure in pounds, k a quantity depending on the figure of the body (equal for a cylinder moving sideways to about 0.77), D the weight of an unit of volume of the fluid (62.3 lbs. for water), v the velocity of the current in feet per second, g the acceleration of gravity, and A the greatest cross-section of the immersed portion of the body, taken at right angles to the direction of motion.

Substituting these values, the formula becomes

$$R = \frac{0.77 \times 62.3}{64.4} Av^2 = 0.745\,Av^2.$$

In order to apply this to the present case, we write $(v - v')^2$ in place of v^2; calling r the radius of the tube, we have $A = \int 2rr'dx$, in which r' is the total depth in feet.

We may now write the following expression for the sum of the forces acting on the tube from N to I:

$$P_1 = 0.745 \int_0^h (v - v')^2\, 2rr'dx.$$

The sum of the moments of these forces, taken with reference to I, is

$$M_1 = 0.745 \times 2rr'^2 \int_0^h (v - v')^2 (h - x)\,dx.$$

We have also the following expression for the sum of the moments of the forces acting on IP, and taken with reference to the same center of moments, I:

$$M_2 = 0.745 \times 2rr'^2 \int_h^l (v' - v)^2 (x - h)\,dx.$$

Performing the indicated operations, and adding the expressions for M_1 and M_2, we have

$$M_1 + M_2 = 1.49\,rr'^2c^2 \Big[-\left(\frac{b^2}{c^2} h^3 + 2\frac{b}{c} h^4 + h^5\right) l + \left(\frac{3}{2}\frac{b^2}{c^2} h^2 + 2\frac{b}{c} h^3 + \frac{h^4}{2}\right) l^2$$
$$-\left(\frac{b^2}{c^2} h - \frac{2}{3} h^3\right) l^3 - \left(\frac{b}{c} h + \frac{h^2}{2} - \frac{b^2}{4c^2}\right) l^4 - \left(\frac{h}{5} - \frac{2}{5}\frac{b}{c}\right) l^5$$
$$+ \frac{b^2}{2c^2} h^4 + \frac{6}{5}\frac{b}{c} h^5 + \frac{11}{15} h^6 + \frac{l^6}{6}\Big].$$

Calling the quantity within the brackets B, we may write

$$M_1 + M_2 = 1.49rr'^2c^2B.$$

If l is the equivalent length, B will be a function of b/c, and the following table gives some of its numerical values:

TABLE 6

$-b/c$	B
0	0.02328
.1	.01836
.2	.01410
.3	.01046
.4	.00736
.5	.00498
.6	.00318
$\frac{2}{3}$	.00231

Since the weight, W, of the tube is the same as the weight of the displaced water, $W = 62.3\pi r^2 r'l$.

The arm $C \sin f$, of the couple, which forms the moment of stability, is the horizontal distance between two vertical lines, one of which is the line drawn through the center of gravity, R, of the tube, and the other the line drawn through the center of gravity of the displaced fluid. Hence $C = (l/2 - RP)r'$, and putting $RP = l/m$, we have $C = (m-2)lr'/2m$. Substituting these values of $M_1 + M_2$, W, and C in the formula $\sin f = (M_1 + M_2)/WC$, we obtain

$$\sin f = \frac{2.98mc^2B}{62.3\pi rl^2(m-2)} = \frac{0.01523mc^2B}{rl^2(m-2)}.$$

Assuming $m = 5$ and $r = 1/12$, which were the proportions adopted in the Lowell Hydraulic Experiments, we have

$$\sin f = \frac{0.3045Bc^2}{L^2}.$$

By means of this formula was computed the value of f given in the table below. It is seen that the deviation from the vertical is very small, and that the tube may be regarded as practically in a vertical position.

The following table gives the results of the application of the preceding principles in determining v, L, and f. The data for the construction of the table were taken from "The Roorkee Hydraulic Experiments" of Capt. Allan

Cunningham, a reference to which was made in article 1. In the work describing these experiments Capt. Cunningham has given an able investigation of the theory of rod-motion, found on pages 240-246. The first column of the following table gives the number of the series of experiments, each being the mean of several trials. The second, third, and fourth columns, give the values of a, b, and c, respectively, as determined by the method of least squares.

The sixth and seventh columns give the values of the equivalent length L, which sometimes has one and sometimes two values, corresponding to the same value of b/c. The eighth column gives the value of f to the nearest minute, which is the deviation of the tube from the perpendicular.

ROORKEE HYDRAULIC EXPERIMENTS.

TABLE 7.

No. of Series	a	b	c	b/c	L		f	No. of Series	a	b	c	b/c	L		f
		+	−	−						+	−	−			
1	4.25	0.07	0.88	0.08	0.943		18′	21	4.44	0.03	0.57	0.05	0.944		8′
2	4.33	.33	1.15	.29	.934		17	22	3.51	.71	1.23	.58	.926		7
3	3.84	.36	1.25	.29	.935		20	23	4.29	.10	.83	.12	.942		14
4	3.48	.49	1.27	.39	.930		15	24	3.41	.26	.89	.29	.935		10
5	4.61	.72	1.11	.65	.925	0	4	25	3.89	.34	.96	.35	.932		10
6	4.27	.65	.99	.66	.925	0	3	26	2.94	.67	1.39	.48	.927		13
7	4.07	.52	.99	.53	.926		5	27	3.34	.30	.85	.35	.932		8
8	4.06	.38	.89	.43	.929		6	28	2.84	− .03	.60	+ .05	.947		
9	4.31	.49	.92	.53	.926		5	29	2.39	+ 1.12	1.01	− 1.11	No equiv. length in this case.		
10	4.50	.16	.55	.29	.935		4	30	2.46	1.49	1.50	.99	.931	0.579	
11	4.05	.51	.97	.53	.926		5	31	2.74	1.09	1.12	.97	.936	.538	
12	3.81	.47	1.20	.39	.930		13	32	3.07	1.43	1.57	.91	.949	.413	
13	4.06	.55	1.04	.53	.926		6	33	3.15	1.34	1.32	1.02	.900	.676	
14	3.89	.84	1.26	.67	.925	0	4	34	3.55	1.64	2.09	.78	.939	.167	
15	4.10	.47	.92	.51	.926		5	35	4.18	.73	1.36	.54	.926		10
16	3.99	.47	1.07	.44	.929		9	36	4.16	.61	1.28	.48	.928		11
17	3.76	.56	1.12	.50	.927		8	37	4.12	.26	.93	.28	.936		12
18	6.43	.43	.72	.60	.926		2	38	3.63	1.24	1.58	.78	.939	.167	
19	6.05	− .33	.11	+ 3.00	.981			39	3.91	.65	1.14	.57	.926		6
20	5.65	+ .16	.85	− .20	.939		12	40	3.15	1.35	1.72	.78	.939	.167	

BOSTON, MASS.,
JULY, 1909.

GENERALIZED GEOMETRIC MEANS AND ALGEBRAIC EQUATIONS

By Otto Dunkel

It has long been known that the arithmetic mean of n positive quantities is greater than or equal to their geometric mean and several proofs of this theorem have been given, one of which is due to Cauchy.* It has also been shown by Hamy that, if all the geometric means taking p of the n quantities at a time be added and divided by the number of means thus obtained, the result will be greater than or equal to the corresponding result taking $p + 1$ at a time; † thus showing that there is a descending series of mean values beginning with the arithmetic and ending with the geometric mean.

The object of this paper is to show that a theorem in regard to the roots of an algebraic equation ‡ leads almost immediately to the above theorem in regard to the relative magnitudes of the arithmetic and geometric means and at the same time establishes that there is another descending series of means intermediate in magnitude between these two and of a form different from those of Hamy.

The inequalities thus obtained may be used to derive sufficient conditions for imaginary roots of an algebraic equation, and a fairly simple set of such conditions will be worked out.

It will also be shown that all of these means can be represented as special values of a function of the n quantities and a continuous variable x, so that this function of x will be a continuous mean. A theorem in regard to this function will be used to derive the inequalities of Hamy and also a few other inequalities existing between the different kinds of means.

A Descending Series of Generalized Geometric Means. We shall prove the following theorem:

---

* Cauchy, *Analyse algébrique* (1821), p. 457.

† M. Hamy, *Bull. des sciences math.*, ser. 2, vol. 14, part 1 (1890), p. 103. If we represent the number of combinations of n things taking p at a time by the symbol ${}_nC_p$, then the inequality above is:

$$\frac{\sum (m_1 m_2 m_3 \ldots m_p)^{\frac{1}{p}}}{{}_nC_p} \geqq \frac{\sum (m_1 m_2 m_3 \ldots m_{p+1})^{\frac{1}{p+1}}}{{}_nC_{p+1}}.$$

A proof of this is given on page 30 inequality (24) of this article.

‡ Dunkel, Annals of Mathematics, ser. 2, vol. 10 (1908), p. 48.

If $m_1, m_2, m_3, \cdots m_n$ are real and positive numbers, then:

$$(1) \qquad \frac{\Sigma m_1}{n} \geqq \left(\frac{\Sigma m_1 m_2}{{}_nC_2}\right)^{\frac{1}{2}} \geqq \left(\frac{\Sigma m_1 m_2 m_3}{{}_nC_3}\right)^{\frac{1}{3}} \geqq \cdots \geqq (m_1 m_2 m_3 \cdots m_n)^{\frac{1}{n}},$$

where $\Sigma m_1 m_2 m_3 \cdots m_i$ means the sum of all the products of the m's taking i of them at a time, and of which there are ${}_nC_i = \dfrac{n!}{i!(n-i)!}$; the equality sign holding for any two means not zero only when $m_1 = m_2 = m_3 = \cdots m_n$.

The n positive numbers $m_1, m_2, m_3, \cdots m_n$ are the roots of an algebraic equation of the nth degree, which it will be convenient to write with binomial coefficients as follows:

$$(2) \quad x^n - na_1x^{n-1} + {}_nC_2a_2x^{n-2} - \cdots + (-1)^i{}_nC_ia_ix^{n-i} + \cdots + (-1)^n a_n = 0,$$

where ${}_nC_ia_i = \Sigma\, m_1m_2m_3 \cdots m_i$.

Let us suppose at first that no m is zero; then no member of the series (1) will be zero and each a_i in (2) will be positive and not zero. Since the roots of the equation (2) are all real, the relation

$$(3) \qquad a_i^2 \geqq a_{i-1}\, a_{i+1}, \qquad i = 1, 2, 3, \cdots n-1,$$

must be satisfied,* and this can be written as the continued inequality

$$(4) \qquad \frac{a_1}{1} \geqq \frac{a_2}{a_1} \geqq \frac{a_3}{a_2} \geqq \cdots \geqq \frac{a_i}{a_{i-1}} \geqq \frac{a_{i+1}}{a_i} \geqq \cdots \geqq \frac{a_{n-1}}{a_{n-2}} \geqq \frac{a_n}{a_{n-1}}.$$

The relations in (4) furnish the following inequalities:

$$(5) \qquad \begin{aligned} \frac{a_1}{1} &\geqq \frac{a_{i+1}}{a_i}, \\ \frac{a_2}{a_1} &\geqq \frac{a_{i+1}}{a_i}, \\ \frac{a_3}{a_2} &\geqq \frac{a_{i+1}}{a_i}, \\ &\;\vdots \\ \frac{a_i}{a_{i-1}} &\geqq \frac{a_{i+1}}{a_i}, \end{aligned}$$

* Dunkel, loc. cit.

and, on taking the product of the left-hand sides and the right-hand sides, we have:

$$(6)\quad a_i \geqq \left(\frac{a_{i+1}}{a_i}\right)^i, \quad \text{or} \quad a_i^{i+1} \geqq a_{i+1}^i, \quad \text{or} \quad a_i^{\frac{1}{i}} \geqq a_{i+1}^{\frac{1}{i+1}}, \qquad i = 1, 2, 3, \cdots n-1.$$

Inserting the value of the a's given in (2) we have the first part of theorem (1) for the case in which no m is zero.

If some of the m's are zero, say $m_{s+1} = m_{s+2} = \cdots m_n = 0$, but no other m is zero; then in (1) all the expressions after the sth are zero, and in (2) the last a not zero is a_s, while in (4) the last ratio that we need consider is $\frac{a_s}{a_{s-1}}$. The reasoning then follows just as before.

If all of the m's are equal it is clear that the equality sign must be used in (1) throughout, and we have just seen that, if some of the m's are zero, the equality sign must be used for all the relations in (1) after a certain one. We shall now show that if two of the expressions in (1) are equal and not zero, then all the m's must be equal. To prove this it will suffice to show that, if not all the m's are equal, $a_1^2 > a_2$, since from this inequality it follows that in the first relation of (4) and of (5) the inequality sign alone must be used, if a_i is not zero, and finally the same thing must be true of (6).

That $a_1^2 > a_2$ can be seen to be true from the fact that, if all the roots of a polynomial are real and at least two are distinct, the same thing must be true of its first derivative; and repeating this reasoning on each derivative in turn we reach the conclusion that the $(n-2)$nd derivative which in this case is $x^2 - 2a_1 x + a_2$, if we drop off a numerical factor, must have real and unequal roots, and therefore the inequality stated above must be true.*

---

*This may also be shown directly as follows:

$$a_1^2 - a_2 = \frac{1}{n^2}\left[(\Sigma m_1)^2 - \frac{2n}{n-1}\Sigma m_1 m_2\right] = \frac{1}{n^2}\left[\Sigma m_1^2 - \frac{2}{n-1}\Sigma m_1 m_2\right] = \frac{1}{n^2(n-1)}\Sigma(m_1 - m_2)^2.$$

Therefore $a_1^2 - a_2$ is greater than zero, if at least two m's are not equal.

This completes the proof of theorem 1.*

Application to Algebraic Equations. It has been shown in a previous paper† that, if any equation written in the form:

$$(7)\qquad x^n + na_1x^{n-1} + {}_nC_2a_2x^{n-2} + \cdots + {}_nC_ia_ix^{n-i} + \cdots + na_{n-1}x + a_n = 0$$

has only real roots, then the following equation:

$$(8)\qquad a_{m-1}x^t + ta_mx^{t-1} + {}_tC_2a_{m+1}x^{t-2} + \cdots + {}_tC_ia_{m+i-1}x^{t-i} + \cdots + ta_{m+t-2}x + a_{m+t-1} = 0$$

has also only real roots, and therefore t real roots, if $a_{m-1} \neq 0$.

Let us suppose that all of the roots of (7) are real and also that $a_{m-1} \neq 0$, then, if we indicate the roots of (8) by $a_1, a_2, a_3, \cdots a_t$, the numbers $a_1^2, a_2^2, a_3^2, \cdots a_t^2$ are all positive, and we can apply to them any one of the inequalities (1), which we have just proven. This can be very readily done since it is a very simple matter to obtain the equation whose roots are the squares of the roots of equation (8);‡ and the coefficients of this

---

* In proving this theorem we made use of the fact that all of the m's and consequently all of the a's were positive; but (3) is true even if some of the m's are negative. Thus for any real quantities:

$$\left[\frac{\Sigma m_1m_2\ldots m_i}{{}_nC_i}\right]^2 - \left[\frac{\Sigma m_1m_2\ldots m_{i-1}}{{}_nC_{i-1}}\right]\left[\frac{\Sigma m_1m_2\ldots m_{i+1}}{{}_nC_{i+1}}\right] \geqq 0.$$

This can also be shown without the use of theorems on the roots of algebraic equations by a direct but somewhat tedious reduction of the left-hand side of this inequality to:

$$\frac{1}{(n-i)i^2}\left(\frac{1}{{}_nC_i}\right)^2\Sigma\left\{(m_1-m_2)^2\left[\Sigma(m_3m_4\ldots m_{i-1})^2 + \frac{1}{{}_{i-1}C_1}\Sigma(m_3m_4\ldots m_{i-2})^2(\Sigma m_{i-1})^2 + \ldots\right.\right.$$
$$\left.\left. + \frac{1}{{}_{i-1}C_t}\Sigma(m_3m_4\ldots m_t)^2(\Sigma m_{t+1}m_{t+2}\ldots m_{i-1})^2 + \ldots\right]\right\}.$$

This expression is always positive if the m's are real. The equality in the preceding footnote is the special case of this in which $i = 1$.

† Dunkel, loc. cit., p. 47. The above statement of the theorem is different from the original but is easily seen to be equivalent to it. From the proof given of the theorem it will be obvious that we might have used the reciprocal equation to (8), i. e., an equation with the coefficients of (8) in reverse order; and this fact will be used in connection with (11).

‡ Cf. Burnside and Panton, *Theory of Equations*, vol. 1, p. 78. If we represent the left-hand side of (8) by $f(x)$, then the equation whose roots are the squares of the roots of (8) may be written:

$$f(\sqrt{z})\cdot f(-\sqrt{z})^{*} = 0.$$

new equation furnish all the expressions such as $\Sigma a_1^2 a_2^2 a_3^2 \cdots a_t^2$ in terms of the a's of equation (7). In this way we should obtain a number of necessary conditions for the reality of all the roots of (7). We shall work out one of the simplest of these sets of conditions, making use of

$$(9) \qquad \frac{a_1^2 + a_2^2 + \cdots + a_t^2}{t} \geqq (a_1^2 a_2^2 \cdots a_t^2)^{\frac{1}{t}},$$

which follows from (1). From the equation of the squared roots of (8), or otherwise, we have:

$$a_1^2 + a_2^2 + \cdots + a_t^2 = \left(\frac{t a_m}{a_{m-1}}\right)^2 - 2\left(\frac{t(t-1)}{2}\,\frac{a_{m+1}}{a_{m-1}}\right)$$

$$= \frac{t^2}{a_{m-1}^2}\left(a_m^2 - \frac{t-1}{t}\,a_{m-1}a_{m+1}\right), \qquad a_1^2 a_2^2 a_3^2 \cdots a_t^2 = \left(\frac{a_{m+t-1}}{a_{m-1}}\right)^2,$$

and inserting these in (9) we have, after division by $\dfrac{t}{a_{m-1}^2}$:

$$(10) \qquad a_m^2 - \frac{t-1}{t}\,a_{m-1}a_{m+1} \geqq \frac{1}{t}\left[a_{m-1}^{1-\frac{1}{t}} a_{m+t-1}^{\frac{1}{t}}\right]^2, \qquad a_{m-1} \neq 0.$$

If we use instead of (8) the equation

$$(11) \quad a_{m+1}x^t + t a_m x^{t-1} + {}_tC_2 a_{m-1} x^{t-2} + \cdots + {}_tC_i a_{m-i+1} x^{t-i} + \cdots + a_{m-t+1} = 0,$$

which must also have t real roots, we shall find by the same reasoning:

$$(12) \qquad a_m^2 - \frac{t-1}{t}\,a_{m-1}a_{m+1} \geqq \frac{1}{t}\left[a_{m+1}^{1-\frac{1}{t}} a_{m-t+1}^{\frac{1}{t}}\right]^2, \qquad a_{m+1} \neq 0.$$

These results may be stated as follows: *If either of the inequalities*

$$(13) \qquad a_m^2 - \frac{t-1}{t}\,a_{m-1}a_{m+1} < \frac{1}{t}\left[a_{m-1}^{1-\frac{1}{t}} a_{m+t-1}^{\frac{1}{t}}\right]^2, \qquad a_{m-1} \neq 0,$$

$$< \frac{1}{t}\left[a_{m+1}^{1-\frac{1}{t}} a_{m-t+1}\right]^2, \qquad a_{m+1} \neq 0,$$

is satisfied for any values of m and t, then the equation (7) *has imaginary roots.*

This test may in special cases indicate the presence of imaginary roots when the test previously given* fails, that is, when $a_m^2 - a_{m-1}\, a_{m+1} \geqq 0$, and so may be regarded as a supplement to that test.

A Continuous Mean. Regarding $m_1, m_2, m_3, \cdots m_n$ as positive numbers no one of which is zero,

$$(14) \qquad y = \left(\frac{m_1^x + m_2^x + \cdots + m_n^x}{n}\right)^{\frac{1}{x}}$$

is a continuous function of the real variable x for all values of x except $x = 0$; and it will be seen later that y can be so defined for $x = 0$ as to be continuous without exception.†

It is clear that for any value of x for which it is defined, y is intermediate in value between the largest and smallest of the m's; also for $x = 1$ it is the arithmetic mean of the m's; for $x = -1$ it is the harmonic mean; and we shall now show that *y increases with x except when all the m's are equal; and that*

$$(15) \qquad \begin{aligned} &\lim_{x=-\infty} y = m_1, && \textit{where } m_1 \textit{ is as small as any other } m, \\ &\lim_{x=0} y = (m_1 m_2 m_3 \cdots m_n)^{\frac{1}{n}}, && \\ &\lim_{x=+\infty} y = m_n, && \textit{where } m_n \textit{ is as large as any other } m. \end{aligned}$$

The values of these limits can be found by putting them in the following form:

$$\text{Lim}\ y = \text{Lim}\ (e^{\log y});$$

so that it suffices to find the limit of $\log y$ in the several cases. Taking first the case of $x = 0$:

$$\lim_{x=0} \log y = \lim_{x=0} \frac{\log(m_1^x + m_2^x + \cdots + m_n^x) - \log(m_1^0 + m_2^0 + \cdots + m_n^0)}{x}$$

$$= \frac{d}{dx} \log(m_1^x + m_2^x + \cdots + m_n^x)\,\big|_{x=0} = \log(m_1 m_2 m_3 \cdots m_n)^{n}.$$

Therefore, $\displaystyle \lim_{x=0} y = (m_1 m_2 m_3 \cdots m_n)^{\frac{1}{n}}.$

---

* Dunkel, loc. cit., p. 48.

† Only real and positive values of m_i^x and of y are considered.

In order to find the limit for $x = -\infty$, we have supposed that m_1 is as small as any other m; then:

$$\operatorname*{Lim}_{x=-\infty} \log y = \log m_1 + \operatorname*{Lim}_{x=-\infty}\left(\frac{1}{x}\log\frac{1+\left(\frac{m_2}{m_1}\right)^x+\cdots+\left(\frac{m_n}{m_1}\right)^x}{n}\right)$$

$$= \log m_1, \qquad\qquad \text{where} \quad \frac{m_i}{m_1} \geqq 1,$$

since the limit on the right in the first line is zero. This completes the proof of the first two limits in (15); and the third limit is found by replacing x by $-x$ and proceeding as before.

It remains now to show that y increases with x; this will be done by showing that

$$(16) \qquad \frac{dy}{dx} = \frac{y}{x}\left[\frac{\Sigma m_i^x \log m_i}{\Sigma m_i^x} - \frac{1}{x}\log\left(\frac{\Sigma m_i^x}{n}\right)\right], \quad x \neq 0,$$

is always positive. Let us suppose then that the value of x is fixed; to simplify the proof we shall write

$$m_i^x = b_i, \qquad \text{or} \qquad m_i = b_i^{\frac{1}{x}},$$

and then (16) becomes:

$$(16') \qquad \frac{dy}{dx} = \frac{y}{x^2}\,\frac{1}{\Sigma b_i}\left[\Sigma b_i \log b_i - \Sigma b_i \cdot \log\left(\frac{\Sigma b_i}{n}\right)\right].$$

Since the part outside of the square brackets is always positive, it is only necessary to show that the expression within is positive. Remembering that $b_1, b_2, b_3, \cdots b_r$ are fixed, we shall examine the following function of t, where t is positive:

$$(17) \qquad u = \sum_1^r b_i \log b_i + t\log t - \left(\sum_1^r b_i + t\right)\log\left(\frac{\sum_1^r b_i + t}{r+1}\right)$$

for a minimum. Its derivative with respect to t is

$$(18) \qquad \frac{du}{dt} = \log t - \log\left(\frac{\sum_1^r b_i + t}{r+1}\right) = \log\left(\frac{r+1}{\frac{1}{t}\sum_1^r b_i + 1}\right),$$

and therefore

$$(18') \qquad \frac{du}{dt} \lesseqqgtr 0 \quad \text{according as} \quad \frac{1}{t}\sum_1^r b_i \gtreqqless r \quad \text{or as} \quad t \lesseqqgtr \frac{1}{r}\sum_1^r b_i;$$

and this tells us that u has for $t = \frac{1}{r}\sum_1^r b_i$ its minimum value, which after reduction is

$$(19) \qquad u_r = \sum_1^r b_i \log b_i - \sum_1^r b_i \log\left(\frac{1}{r}\sum_1^r b_i\right).$$

We have then, putting $t = b_{r+1}$ in (17):

$$(20) \qquad u_{r+1} \geqq u_r \quad \text{according as} \quad b_{r+1} \mathrel{\substack{\neq \\ =}} \frac{1}{r}\sum_1^r b_i .$$

Giving r in turn the values 1, 2, 3, $\cdots$ n, and noting that $u_1 = 0$, we have:

$$(20') \qquad u_n \geqq u_{n-1} \geqq u_{n-2} \geqq \ldots \geqq u_3 \geqq u_2 \geqq u_1 = 0.$$

If there are two b's not equal, we may consider them as b_1 and b_2; and from (20) we see that in this case $u_2 > u_1 = 0$, and therefore u_n is greater than zero. Since u_n is the expression in the brackets of (16'), this completes the proof that, if not all the b's are equal, i. e., if not all the m's are equal, $\frac{dy}{dx}$ is always positive. This also furnishes the proof of (15).

If then for $x = 0$ we assign to y the value of the geometric mean of all the m's, it is easily seen that this completed definition makes *y a continuous and increasing function for all values of x.*

The case in which some of the m's are zero can now be easily treated; for, if we suppose that there are $n + k$ of the m's of which the first n are each different from zero but $m_{n+1} = m_{n+2} = \cdots = m_{n+k} = 0$, then (14) may be written:

$$(21) \qquad y = \left(\frac{m_1^x + m_2^x + \cdots + m_n^x}{n}\right)^{\frac{1}{x}} \left(\frac{n}{n+k}\right)^{\frac{1}{x}}, \quad x > 0,$$

and we may define y as zero when $x \leqq 0$.

The first factor in the first expression for y has already been shown to increase with x, and it is easily seen that the second factor also increases with

x and approaches unity as x becomes infinite through positive values, but approaches zero as x approaches zero through positive values. Hence in this case y increases with x for positive values of x and approaches the largest m as x increases indefinitely.

From what preceeds it will be seen that if we define a function of x as follows:

$$(22) \qquad y = \left(\frac{m_1^x + m_2^x + \cdots + m_n^x}{p}\right)^{\frac{1}{x}},$$

where none of the m's are zero and $p \neq n$, this function of x will be discontinuous for $x = 0$ owing to the discontinuity of $\left(\frac{n}{p}\right)^{\frac{1}{x}}$ at this point.

Derivation of Inequalities between Several Forms of Means. In what follows we shall suppose that none of the m's in the expression (14) for y are zero; and in this case, if x and y were plotted, we would obtain a continuous curve rising constantly from the least m at $-\infty$ to the greatest m at $+\infty$ and passing through the harmonic mean at -1, the geometric mean at 0, and the arithmetic mean at $+1$. Also y must take on each and every value between the least and greatest m once and only once. It follows then that each of the means which have been mentioned here, and indeed any mean whatsoever of these n numbers, must be represented by a single point on this curve. After proving the inequalities of Hamy, * we shall derive certain other inequalities which give some idea of the order in which the different means are located on the curve.

In order to prove that:

$$(23) \qquad \frac{\Sigma m_1}{n} \geqq \frac{\Sigma(m_1 m_2)^{\frac{1}{2}}}{{}_nC_2} \geqq \frac{\Sigma(m_1 m_2 m_3)^{\frac{1}{3}}}{{}_nC_3} \geqq \cdots \geqq (m_1 m_2 m_3 \cdots m_n)^{\frac{1}{n}}$$

we shall apply (15) which tells us that

$$\left[\frac{\Sigma(m_1 m_2 m_3 \cdots m_p)^{\frac{1}{p}}}{{}_nC_p}\right]^p \geqq \left[\frac{\Sigma(m_1 m_2 m_3 \cdots m_p)^{\frac{1}{p+1}}}{{}_nC_p}\right]^{p+1};$$

but from (1) we have, after replacing m_i by $m_i^{\frac{1}{p+1}}$,

$$\left[\frac{\Sigma(m_1 m_2 m_3 \cdots m_p)^{\frac{1}{p+1}}}{{}_nC_p}\right]^{\frac{1}{p}} \geqq \left[\frac{\Sigma(m_1 m_2 m_3 \cdots m_{p+1})^{\frac{1}{p+1}}}{{}_nC_{p+1}}\right]^{\frac{1}{p+1}}.$$

* Loc. cit.

From these two inequalities we have

$$(24) \qquad \frac{\Sigma(m_1m_2m_3\cdots m_p)^{\frac{1}{p}}}{{}_nC_p} \geqq \left[\frac{\Sigma(m_1m_2m_3\cdots m_p)^{\frac{1}{p+1}}}{{}_nC_p}\right]^{\frac{p+1}{p}} \geqq \frac{\Sigma(m_1m_2m_3\cdots m_{p+1})^{\frac{1}{p+1}}}{{}_nC_{p+1}},$$

and this gives the relations in (23).

From (1) it follows on replacing m_i by $m_i^{\frac{1}{p}}$ that:

$$(25) \qquad \left[\frac{m_1^{\frac{1}{p}} + m_2^{\frac{1}{p}} + m_3^{\frac{1}{p}} + \cdots + m_n^{\frac{1}{p}}}{n}\right]^p \geqq \frac{(\Sigma m_1m_2m_3\cdots m_p)^{\frac{1}{p}}}{{}_nC_p}.$$

Corresponding members of (1) and (23) may be compared as follows:

$$(26) \qquad \left[\frac{\Sigma m_1m_2m_3\cdots m_p}{{}_nC_p}\right]^{\frac{1}{p}} = \left[\frac{\Sigma\left[(m_1m_2m_3\cdots m_p)^{\frac{1}{p}}\right]^p}{{}_nC_p}\right]^{\frac{1}{p}} \geqq \frac{\Sigma(m_1m_2m_3\cdots m_p)^{\frac{1}{p}}}{{}_nC_p},$$

the inequality in the above following from (15), since $p \geqq 1$.

Thus each member of (1) is greater than or equal to the corresponding member of (23).

It would be interesting to have other inequalities corresponding to (25) which would show in what intervals on the x axis each member of (1) and (23) would appear when plotted on the curve of (14).

Approximations to the Roots of Algebraic Equations. If it is known that all the roots of an algebraic equation are real, the properties of the function y in (14) may be used to obtain approximations to their absolute values. Let us suppose that the roots arranged in order of their numerical magnitude are $m_1, m_2, m_3, \cdots m_n$, so that m_n is as large in absolute value as any other root. If now x takes on only even integral values, the results in (15) will still be true with a slight modification, whether all of the roots are positive or not, and this modification consists in replacing the m's by their absolute values in the values of the different limits; thus two of the limits in (15) would now read:

$$\lim_{x=-\infty} y = |m_1| ; \qquad \lim_{x=+\infty} y = |m_n| .$$

If then to x is assigned the series of values 2, 4, 6, 8, $\cdots$, there will result a set of increasing values:

$$y_2 < y_4 < y_8 < \cdots,$$

which will approach $|m_n|$ as a limit. On the other hand, corresponding to the values -2, -4, -6, -8, $\cdots$ of x, there is a descending series:

$$y_{-2} > y_{-4} > y_{-6} > y_{-8} > \cdots$$

approaching $|m_1|$ as a limit.

Considering next the following function of x:

$$\left[\frac{\Sigma(m_1m_2)^x}{{}_nC_2}\right]^{\frac{1}{x}}, \tag{27}$$

we would obtain in the same way a set of increasing values corresponding to increasing values of x and approaching $|m_{n-1}|\,|m_n|$ as a limit. Taking the quotient of (14) by (27) we obtain a function of x:

$$\left[\frac{\Sigma(m_1m_2)^x}{\Sigma m_1^x}\,\frac{n}{{}_nC_2}\right]^{\frac{1}{x}},$$

which will approach $|m_{n-1}|$ as a limit; and in general

$$\left[\frac{\Sigma(m_1m_2\ldots m_p)^x}{\Sigma(m_1m_2\ldots m_{p-1})^x}\,\frac{p}{n-p+1}\right]^{\frac{1}{x}} \tag{28}$$

will approach $|m_{n-p+1}|$ as a limit as x increases indefinitely.

The numerical values of the expressions $\Sigma(m_1\, m_2 \cdots m_p)^x$ in the successive approximations (28) can be obtained most easily, if we consider only those values of x which are powers of 2, by forming the equation whose roots are the squares of the roots of the original equation;* then the equation whose roots are the squares of the roots of the equation just found and so on: so that if the tth equation thus found is:

$$z^n + C_1^{(t)} z^{n-1} + C_2^{(t)} z^{n-2} + \ldots + C_{n-1}^{(t)} z + C_n^{(t)} = 0,$$

---

* Cf. footnote ‡ page 24.

then its roots are $m_1^{2^i}, m_2^{2^i}, m_3^{2^i} \cdots$ and the numerical value of $\Sigma(m_1 m_2 m_3 \cdots m_p)^{2^i}$ is $(-1)^p C_p^{(i)}$. Inserting these values in (28) we have:

$$\left[-\frac{C_p^{(i)}}{C_{p-1}^{(i)}} \frac{p}{n-p+1}\right]^{\frac{1}{2^i}} \tag{29}$$

as an approximation to $|m_{n-p+1}|$. Forming all the n expressions (29) corresponding to the given equation, all of its roots may in this way be approximated simultaneously.*

COLUMBIA, MISSOURI,
APRIL, 1909.

---

*Cf. Netto, *Algebra*, vol. 1, p. 290–297. On page 292 the approximations to the three roots of a cubic are worked out. The expressions in Netto corresponding to (29) above do not contain the factor $\frac{p}{n-p+1}$.

THE GEOMETRY OF CHAINS ON A COMPLEX LINE*

By John Wesley Young

Introduction. This paper forms an elementary chapter in the projective geometry on a complex line, i. e., of a line whose points are isomorphic with the system of ordinary complex numbers and infinity. It furnishes a synthetic treatment of certain well-known topics in the theory of functions of a complex variable; it has close contact with recent work on the foundations of projective geometry; finally, it forms the basis for certain generalizations, which offer a powerful synthetic approach to various analytic problems. To this last aspect of the paper I shall return in detail on a future occasion.† The first two aspects I discuss briefly in this introduction.

But first, in view of the fact that the notion of a chain of points on a line does not appear to be generally familiar, it seems desirable to describe it briefly. This is most readily done in analytic language. Any point of a projective line may be conveniently represented by a single coordinate x. If the points of a line are hereby brought into one-to-one correspondence with the ordinary real numbers and infinity, we are wont to speak of a real line; if on the other hand the points of a line are hereby made isomorphic with the system of ordinary complex numbers and infinity we speak of a complex line. A *real representation* of the points of a complex line may then be obtained by the usual method for representing complex numbers by the real points of an Argand plane. The points of the line with real coordinate x form a subset $\mathfrak{C}_0$ of points on the line:— in the representation mentioned this subset $\mathfrak{C}_0$ is represented by the points of the real axis. By means of a projective transformation on the line the set $\mathfrak{C}_0$ is transformed into another or the same set $\mathfrak{C}$. Any such set of points $\mathfrak{C}$ on the line which is projective with the set $\mathfrak{C}_0$ of real points on the line is called a *chain* on the line. In the Argand representation the chains on the line evidently correspond to the circles (and straight lines) of the plane.

---

* Presented to the American Mathematical Society February 29, 1908.

† Cf., however, the last paragraph of the present paper.

The contact of this paper with the theory of functions of a complex variable is now evident. The study synthetically of the chains on a line with reference to their behavior toward the projectivities on the line and the resulting classification of the projectivities (into elliptic, hyperbolic, parabolic, and loxodromic) furnishes a synthetic derivation of well-known theorems concerning the linear fractional substitutions on a complex variable. On account of its elementary character this method of treatment may be of pedagogic value.*

The notion of a chain has been fundamental in the synthetic introduction of imaginaries into geometry since the time of von Staudt.† In the more recent work on the foundations of projective geometry it necessarily plays an important role. Pieri has indeed recently chosen the chain as one of the undefined elements in his set of assumptions for complex projective geometry.‡ More recently Professor Veblen and I § have given a set of assumptions for projective geometry in which the point and an undefined class of points called a line are the only undefined elements, and in which the chain is defined. This paper will in the sequel be referred to occasionally by the letter A. In its relation to the foundations the present paper forms a continuation of the one just referred to, in which it is shown how the fundamental properties of chains are derived from the assumptions laid down in A. As may be expected several of Pieri's assumptions occur here as theorems. From the point of view of the foundations it is interesting to note that the projective geometry of chains on a complex line is isomorphic with the metric geometry of circles in a plane.

From what has been said it is clear that the approach to the present paper may be either from the synthetic or from the analytic side. For the former reference may simply be made to A. The analytic description of a chain is given above. The presuppositions of the paper are given below in section 1. They are for the most part elementary notions and theorems of projective

* If the isomorphism between the system of complex numbers and the points of a complex projective line be emphasized together with the fact that the usual Argand representation is simply a *real representation of the points of a complex line*, the beginner will no longer be troubled with the fact that the Argand plane shows only a point at infinity, whereas the projective plane has a line at infinity.

† v. Staudt, *Beiträge zur Geometrie der Lage*, Nürnberg, 1857, p. 137 ff.

‡ Pieri, Nuovi principii di geometria projettiva complessa, *Memorie della R. Accad. d. Scienze di Torino*, ser. 2, vol. 55 (1905), pp. 189–235.

§ Veblen and Young, A set of assumptions for projective geometry, *American Journal of Mathematics*, vol. 30 (1908), pp. 347–380.

geometry; in so far as they apply to the less familiar ideas of complex projective geometry reference may be made to A, or they may be derived analytically without difficulty. In section 2 is discussed the system of invariant chains of an involution. Conjugate points with respect to a chain and the notion of orthogonal chains are defined and discussed in section 3. The following section brings the classification of projectivities into elliptic, hyperbolic, parabolic, and loxodromic, and the discussion of the associated systems of chains. In section 5 are then derived certain analytic criteria. Finally in section 6 the results previously obtained are used to write down a complete list of the different types of continuous groups which leave a chain invariant, i. e., of those continuous groups of linear fractional transformations which can be represented with real coefficients. This section and the paper close with a brief reference to the application of the results and methods of this paper to the synthetic treatment of the geometric theory of discontinuous groups of projectivities, i. e., of the groups of the automorphic functions of one variable.

SECTION 1. **Presuppositions and notation.** Points will in general be denoted by the capital letters of the alphabet such as A, B, $\cdots$ M, N. All points considered are on the same complex line. We assume the following definitions and theorems from projective geometry:

I. The *harmonic conjugate* B of a point A with respect to two others CD is a uniquely determined point, distinct from A, C, and D; this is denoted by $B = H(A, CD)$. The relation $B = H(A, CD)$ implies the relations $A = H(B, CD)$ and $C = H(D, AB)$. Each of the pairs AB, CD, is said to be *harmonic* with the other; and the two together form a *harmonic set*; this is denoted by $H(AB, CD)$.

II. 1. A *projectivity* π is a one-to-one reciprocal correspondence or transformation, whereby to any point P corresponds a unique point P'; in symbols $\pi(P) = P'$. 2. A projectivity transforms any harmonic set into a harmonic set.* 3. A projectivity is uniquely determined by three pairs of homologous points. 4. A point M for which $\pi(M) = M$ is called a double point of π; a projectivity is uniquely determined when two double points and one pair of homologous points is given. 5. If a projectivity π_1 is followed by another or the same π_2, the resultant correspondence is a projectivity denoted by $\pi_2\pi_1$; the inverse of a given projectivity π is denoted by π^{-1}. 6. Every projectivity different from the identity has at least one and not

*On a complex line properties II 1, 2 are not sufficient to characterize a projective transformation. Cf. A, p. 373.

more than two double points. 7. Any two projectivities with the same double points are commutative, i.e., $\pi_2\pi_1 = \pi_1\pi_2$. 8. If π and $\pi_1\pi\pi_1^{-1}$ have the same double points we have $\pi_1\pi\,\pi_1^{-1} = \pi$. 9. If a projectivity π has only one double point M and we have $\pi(AA_1) = A_1A_2$, we have $H(AA_2, A_1M)$.

III. 1. A projectivity I of period two, i.e., such that $II = 1$, is called an *involution*. 2. Any involution has two distinct double points, and 3. is determined uniquely by two pairs of homologous points, in particular by its double points; the involution I determined by the two homologous pairs AA', BB' is denoted by $I(AA', BB')$. 4. If the double points of an involution I are MN, and if AA' is any homologous pair, we have $H(AA', MN)$, and conversely; a homologous pair of an involution is ususually called a conjugate pair. 5. If a projectivity interchanges two distinct points it is an involution.

IV. 1. A *chain* $\mathfrak{C} = |ABC|$ is an infinite class of points, uniquely determined by any distinct three of its points ABC. 2. Any class of points projective with a chain is a chain. 3. If PQR are any three distinct points of a chain, $S \doteq H(P, QR)$ is a point of the chain.

V. 1. Neither double point of the involution $I(AA', BB')$ determined by a harmonic set $H(AA', BB')$ is a point of $|AA'B|$. 2. Every chain containing two conjugate pairs of an involution I but not containing the double points of I has at least one point in common with every chain through the double points. 3. Through a point P of a chain $\mathfrak{C}$ and any point Q not on $\mathfrak{C}$ there is not more than one chain that has no other point in common with $\mathfrak{C}$ than P.

The last two propositions (V, 2, 3) merit a word of comment. From the point of view of the foundations, each of them is a consequence of the other in connection with the other assumptions made in A for complex projective geometry; one of them, however, is necessary as an assumption of closure. In A we chose V, 3 for this purpose, which is there given as Assumption I 2 (A, p. 371.) The reader not interested in the foundations may simply assume these propositions as intuitionally evident from the Argand representation; chosing $H(AA', BB') = H(0\,\infty,\,-1\,1)$, for example, in which case the double points of I in V, 2 are i and $-i$.

SECTION 2. **The invariant chains of an involution.** Throughout this section I represents an involution whose double points are MN.

THEOREM 1. *An involution I leaves invariant every chain containing the double points M, N of I.*

Proof. Let $\mathfrak{C}$ be any chain containing the double points and let A be any point of $\mathfrak{C} = |MNA|$. Then $I(A) = A' = H(A, MN)$ (by I, 1, and III, 4). Hence A' is a point of $|MNA|$ by IV, 1, 3.

THEOREM 2. *Through every point $P(\neq M, N)$ there is an invariant chain of I, not containing MN.*

Proof. Let P' be the conjugate of P in the involution I, and let I_1 be $I(MN, PP')$. The double points QQ' of I_1 are harmonic with MN, and hence form a conjugate pair of I (III, 2, 3, 4). But we also have $H(PP', QQ')$ (III, 4). Hence the chain $|PQP'| = |P'Q'P|$ is invariant under I, and does not contain the double points MN of I (V, 1).

THEOREM 3. *An invariant chain of I not through MN is cut by every chain through MN in two points harmonic with MN.*

Proof. Let $\mathfrak{C}$ be an invariant chain not through MN. Then, by V, 2, every chain through MN cuts $\mathfrak{C}$ in at least one point, which is not a double point of I. Since every chain through MN is also invariant by Theorem 1, it follows that every such chain cuts $\mathfrak{C}$ in two distinct points which are conjugates of I, and are therefore harmonic with MN (III, 4).

THEOREM 4. *If a chain $\mathfrak{C}$ not through MN meets two chains through MN in pairs of points harmonic with MN, it meets every chain through MN in points harmonic with MN.*

Proof. Let $\mathfrak{C}_1$, $\mathfrak{C}_2$ be the two chains through MN and let A_1B_1 and A_2B_2 be the pairs of points in which they meet $\mathfrak{C}$ respectively. Then MN are the double points of $I(A_1B_1, A_2B_2)$, and this involution leaves $\mathfrak{C} = |A_1B_1A_2| = |B_1A_1B_2|$ invariant. The theorem then follows from Theorem 3.

DEFINITION. A chain $\mathfrak{C}$ which cuts every chain through two given points MN in points harmonic with MN is said to be *about** MN. The two points MN are said to be *conjugate* with respect to $\mathfrak{C}$, and each is the *conjugate* or *inverse* of the other with respect to $\mathfrak{C}$; every point of a chain $\mathfrak{C}$ is said to be *conjugate with* or *inverse to* itself with respect to $\mathfrak{C}$.

THEOREM 5. *Through any point $P(\neq M, N)$ there is one and only one chain about MN.*

* "About" is here used in a technical sense. Cf. Harkness and Morley, *Introduction to the theory of analytic functions* (London, 1898), p. 30.

Proof. By Theorems 2 and 3, there is one chain $\mathfrak{C}$ through P about MN. Moreover, using the notation in the proof of Theorem 2, we can show that $\mathfrak{C} = |PQP'|$ is the only chain through P about MN. For, if $\mathfrak{C}'$ were another chain through P about MN, $\mathfrak{C}'$ would also contain P' (fig. 1). Now, $|QMN|$ is about PP' by Theorem 4 and hence $\mathfrak{C}$ and $\mathfrak{C}'$ meet $|QMN|$ in two

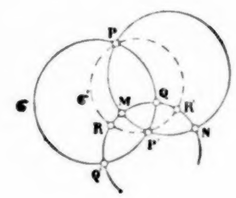


Fig. 1.

pairs of points QQ', RR' each of which is harmonic with PP'. But by hypothesis the pairs QQ', RR' are harmonic with MN. This is impossible, since MN and PP' cannot both be the double points of $I(QQ', RR')$ (III, 4).

Corollary 1. *The totality of chains about two given points MN forms an infinite system of chains no two of which have a point in common.*

For, only one chain of the system can pass through any given point, and there is one chain of the system through every point ($\neq M, N$) of any chain through MN.

Corollary 2. *The chains through two points form an infinite system of chains.*

For there is one through every point on any chain about the two points.

Theorem 6. *An involution whose double points are MN leaves invariant every chain through MN and every chain about MN, and only these.*

Proof. Clearly an invariant chain of an involution must contain both or neither of the double points (III, 4). An involution leaves invariant every chain through the double points by Theorem 1. Every chain about MN meets every chain through MN in a pair of points harmonic with MN. Hence, any point of a chain about MN is conjugate with a point of the same chain (III, 4). Further, every invariant chain not through MN must meet every

chain through MN in points harmonic with MN (Theorem 3), and is therefore about MN (III, 4).

Corollary. *Through every point $P(\neq M, N)$ pass two and only two invariant chains of the involution whose double points are MN.*

It should be here noted that any two distinct points MN define two infinite systems of chains, viz., the system through MN and the system about MN, such that through every point distinct from M and N passes one and only one chain of each system and such that every chain of one system meets each chain of the other in points harmonic with MN. These two systems are of great importance in the geometry of chains on a line.

Theorem 7. *If a projectivity π with distinct double points MN leaves a chain through MN invariant, it leaves every chain through MN invariant.*

Proof. Let $\mathfrak{C}$ be the given invariant chain through MN, and let $A(\neq M, N)$ be any point of $\mathfrak{C}$, and $B(\neq M, N)$ any point of any other chain $\mathfrak{C}'$ through MN. Then if π_1 is determined by $\pi_1(MNA) = MNB$, we have $\pi_1(\mathfrak{C}) = \mathfrak{C}'$. But $\pi_1\pi\pi_1^{-1}$ evidently leaves $\mathfrak{C}'$ invariant, and $\pi_1\pi\pi_1^{-1} = \pi$, since π, π_1 have the same double points (II, 7). Hence, π leaves $\mathfrak{C}'$ invariant.

Section 3. Conjugate points and orthogonal chains.

Theorem 8. *If a chain, $\mathfrak{C}$, is about two points MN, and π is any projectivity, then $\pi(\mathfrak{C})$ is about $\pi(M)\ \pi(N)$.*

Proof. If any chain, $\mathfrak{C}_1$, through $\pi(M)\ \pi(N)$ failed to cut $\pi(\mathfrak{C})$ in points harmonic with $\pi(M)\ \pi(N)$, $\pi^{-1}(\mathfrak{C}_1)$ would be a chain through MN not cutting $\mathfrak{C}$ in points harmonic with MN.

Corollary 1. *If MN are conjugate with respect to $\mathfrak{C}$, $\pi(M)\ \pi(N)$ are conjugate with respect to $\pi(\mathfrak{C})$.*

Corollary 2. *Every projectivity with distinct double points MN leaves invariant each of the systems of chains through MN and about MN.*

Definition. Two chains are said to *meet orthogonally* or to be *orthogonal*, if one contains two points which are conjugate with respect to the other.

Theorem 9. *Every projectivity transforms a pair of orthogonal chains into a pair that is orthogonal.*

This is an immediate consequence of Theorem 8.

Theorem 10. *If $\mathfrak{C}$ and $\mathfrak{C}'$ are two chains meeting orthogonally in PQ,*

then every pair of points MN on $\mathfrak{C}(\mathfrak{C}')$ which are harmonic with PQ are conjugate with respect to $\mathfrak{C}'$ $(\mathfrak{C})$.

Proof. Let AB on $\mathfrak{C}$ be conjugate with respect to $\mathfrak{C}'$, so that we have $H(AB, PQ)$ (fig. 2). Let $A'B'$ be the double points of $I(AB, PQ)$. $A'B'$ are on $\mathfrak{C}'$. For $I(PQ, A'B')$ leaves $|A'PB'|$ invariant and this chain must

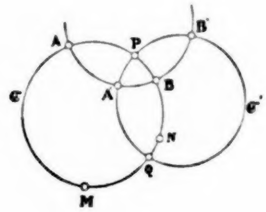


FIG. 2.

therefore be about AB: but there is only one chain through P and about AB. Then $\mathfrak{C}$ cuts $\mathfrak{C}'$ and $|AA'B|$ in pairs of points harmonic with $A'B'$. Hence $A'B'$ are conjugate with respect to $\mathfrak{C}$: i. e., *each of two orthogonal chains contains a pair of points conjugate with respect to the other.* Now, as before, let AB be conjugate with respect to $\mathfrak{C}'$, and let $\pi(PQAB) = PQMN$, which is possible since AB and MN are each harmonic with PQ. π transforms $\mathfrak{C}$ into itself since $|PQA| = |PQM|$ and therefore also leaves $\mathfrak{C}'$ invariant (Theorem 7). Hence, MN are conjugate with respect to $\mathfrak{C}'$ (Theorem 8, Corollary 1).

COROLLARY. *Through any point of either of two orthogonal chains which is not on both passes a chain orthogonal to both.*

This is the chain through the given point and about the points common to the two given chains.

THEOREM 11. *The inverse of any point P with respect to a given chain $\mathfrak{C}$ is a uniquely determined point.*

Proof. If P is on $\mathfrak{C}$ the theorem is immediate from the definition. If P is not on $\mathfrak{C}$, let AB be any pair of points conjugate with respect to $\mathfrak{C}$.*

* Such a pair exists; e. g., the double points of $I(CC', DD')$, where $H(CC', DD')$ is any harmonic set on $\mathfrak{C}$.

The chain $|ABP|$, orthogonal to $\mathfrak{C}$, meets $\mathfrak{C}$ in two points MN. Hence $Q = H(P, MN)$ is the inverse of P with respect to $\mathfrak{C}$ (Theorem 10). This point Q is unique, for if there were another Q', $|PQQ'|$ would cut $\mathfrak{C}'$ in two points harmonic with PQ and with PQ' which is absurd (I, 1).

THEOREM 12. *Through any two distinct points AB there is one and only one chain orthogonal to a given chain $\mathfrak{C}$.*

Proof. Suppose A is not on $\mathfrak{C}$, and let A' be the inverse of A with respect to $\mathfrak{C}$. Then $|AA'B|$ is orthogonal to $\mathfrak{C}$ and it is the only chain through AB with this property, since any such chain must contain A'. This applies whether B is on $\mathfrak{C}$ or not. If both A and B are on $\mathfrak{C}$, let MN be any pair on $\mathfrak{C}$ harmonic with AB. Then the chain through A about MN contains B. This chain then satisfies the conditions of the theorem. It is the only one, since every such chain must be through A and about MN and there is only one such (Theorem 5).

SECTION 4. **Hyperbolic, elliptic and parabolic projectivities.**

THEOREM 13. *If a non-involutoric projectivity, π, leaves invariant a chain $\mathfrak{C}$, then π leaves invariant one and only one chain through every point which does not coincide with a double point of π.*

Proof. 1) Let π have two distinct double points MN (II, 6). Let A $(\neq M, N)$ be any point of $\mathfrak{C}$, and B any other point distinct from MN. Let π_1 be determined by $\pi_1(MNA) = MNB$. $\pi(\mathfrak{C}) = \mathfrak{C}'$ is then a chain through B. Since $\pi_1\pi\pi_1^{-1}$ clearly leaves $\mathfrak{C}'$ invariant and $\pi_1\pi\pi_1^{-1} = \pi$ (II, 7), it follows that π leaves $\mathfrak{C}'$ invariant. Moreover, suppose $\mathfrak{C}'_1$, were another chain through B which is left invariant by π. $\mathfrak{C}'_1$ has another point B' in common with $\mathfrak{C}'$, since B is not a double point. π must then simply interchange BB', in which case π would be involutoric (III, 5), contrary to hypothesis. Hence $\mathfrak{C}'$ is the only invariant chain through B.

2) Let π have only one double point M(II, 6). Let $A(\neq M)$ be any point of $\mathfrak{C}$, and B $(\neq M)$ any other point. Let $\pi_1(MMA) = MMB$; the theorem then follows as before.

COROLLARY 1. *If a projectivity leaves two distinct chains through a point distinct from a double point invariant, it is an involution.*

COROLLARY 2. *The system of invariant chains of a non-involutoric projectivity with distinct double points consists of all the chains* $\substack{\text{through}\\ \text{about}}$ *the double points, if one invariant chain is* $\substack{\text{through}\\ \text{about}}$ *them.*

For, if $\mathfrak{C}$ is $\substack{\text{through}\\\text{about}}$ MN, the projectivity π_1 transforms $\mathfrak{C}$ into another chain $\substack{\text{through}\\\text{about}}$ MN (Theorem 8); and every chain $\substack{\text{through}\\\text{about}}$ MN can be obtained in this way.

Definition. A projectivity with two distinct double points which leaves a chain invariant, is called *hyperbolic* or *elliptic* according as an invariant chain does or does not contain the double points*. A projectivity with two double points which does not leave any chain invariant is called *loxodromic*. A projectivity with only one double point is called *parabolic*.

This classification is mutually exclusive, except that an involution is both elliptic and hyperbolic. Moreover, *every projectivity belongs to one of these four classes*. To show this we need only show that, *if a projectivity* π, *with two double points leaves invariant a chain containing one double point this chain also contains the other*.

To prove this let MN be the double points of π and let $\mathfrak{C}$ be an invariant chain of π containing M. Let P ($\neq M, N$) be any other point of $\mathfrak{C}$ and let I be the involution in which MN are conjugate and in which P is a double point. Then we have $I\pi I = \pi$, since $I\pi I$ has the same double points as π (II, 8). I transforms $\mathfrak{C}$ into a chain $\mathfrak{C}'$ through N, which is invariant under $I\pi I = \pi$. $\mathfrak{C}$ and $\mathfrak{C}'$ are both invariant chains of π and both contain P. Hence either $\mathfrak{C} = \mathfrak{C}'$, or π is an involution (Theorem 13).

Theorem 14. *Every hyperbolic projectivity leaves every chain through the double points invariant and, unless it is involutoric, only these. Every elliptic projectivity leaves every chain about the double points invariant and, unless it is involutoric, only these.*

Proof. The first part of this theorem is contained in Theorem 13, Corollary 2. The second part will likewise follow from this corollary, if we show that every elliptic projectivity, π, must leave one chain about the double points invariant. Let $\mathfrak{C}$ be invariant under π; it does not then contain either of the double points MN. We note first that there exists a chain, $\mathfrak{C}'$, through the double points MN meeting $\mathfrak{C}$ in two distinct points. For let M' be the inverse of N with respect to $\mathfrak{C}$. If we have $M' = M$, our theorem is proved. Suppose $M' \neq M$; the chain $|M'MN|$ meets $\mathfrak{C}$ in two points AB. Now, let $\pi(AA_1BB_1) = A_1A_2B_1B_2$; A_1, A_2, B_1, B_2 are then on $\mathfrak{C}$. Further let π_1

* This definition of elliptic and hyperbolic projectivities seems the most natural and agrees with the definition given by some writers; others make the class of hyperbolic substitutions less extensive. Cf. footnote page 46.

be determined by $\pi_1(MNA) = MNB$. Then we have $\pi_1^{-1}(B_1) = \pi_1^{-1}\pi\pi_1(A) = \pi(A) = A_1$, since π, π_1 are commutative (II, 7). Hence $\pi_1(A_1) = B_1$. Similarly, we obtain $\pi_1(AA_1A_2) = BB_1B_2$. Hence π_1 leaves $\mathfrak{C}$ invariant. Also, π_1 leaves $\mathfrak{C}' = |MNA| = |MNB|$ invariant. Hence, π_1 is an involution (Theorem 13, Corollary 1), and therefore $\mathfrak{C}$ is about MN. (Theorem 6).

COROLLARY. *Conversely, to every system of chains* $\substack{\textit{through}\\ \textit{about}}$ *two points* MN *corresponds an infinite set of* $\substack{\textit{hyperbolic}\\ \textit{elliptic}}$ *projectivities, each of which leaves invariant every chain* $\substack{\textit{through}\\ \textit{about}}$ MN, *and permutes among themselves the chains* $\substack{\textit{about}\\ \textit{through}}$ MN.

For, if AB are any two points of a chain through MN, the projectivity π determined by $\pi(MNA) = MNB$, leaves each chain of the system through MN invariant; and if AB are any two points of a chain about MN the projectivity determined in the same way will leave every chain of the system about MN invariant.

Loxodromic projectivities, moreover, exist. For, if in the above AB are two points not on the same chain through MN nor on the same chain about MN, the projectivity there described will necessarily be loxodromic.

THEOREM 15. *Every parabolic projectivity π with double point M leaves invariant each chain of a system, any two of which have M and no other point in common. Through any point $A (\neq M)$ passes one and only one chain of this system.*

Proof. Let $\pi\,(AA_1) = A_1A_2$; then we have $H(AA_2, A_1M)$ (II, 9). π therefore leaves $|MAA_1| = |MA_1A_2|$ invariant. This also shows that every invariant chain contains M. Finally, two invariant chains cannot have any point other than M in common, since such a point would have to be a double point (or by Theorem 13).

DEFINITION. Two chains with a point M and no other point in common are said to be *tangent* or to *touch at* M; the point M is called the *point of tangency* or *of contact*.

We have just seen that the invariant chains of a parabolic projectivity with double point M form a system of chains mutually tangent at M.

THEOREM 16. *Any chain $\mathfrak{C}$ through the double point M of a parabolic projectivity, π, is transformed by π into another tangent to $\mathfrak{C}$ at M unless $\mathfrak{C}$ is an invariant chain of π.*

Proof. Suppose $\pi\ (\mathfrak{C}) = \mathfrak{C}'$ has another point in common with $\mathfrak{C}$; let it be denoted by A_1. Let A_0 and A_2 be determined by $\pi(A_0 A_1) = A_1 A_2$. Then clearly $\mathfrak{C} = |A_0 A_1 M|$ and $\mathfrak{C}' = |A_1 A_2 M|$. But we have $H(A_0 A_2, A_1 M)$, whence would follow $\mathfrak{C} = \mathfrak{C}'$.

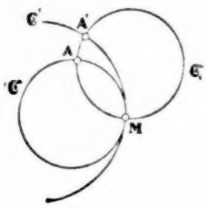


FIG. 3.

THEOREM 17. *If two chains $\mathfrak{C}$ and $\mathfrak{C}_1$ meet orthogonally in a point M, every chain tangent to $\mathfrak{C}$ at M meets every chain tangent to $\mathfrak{C}_1$ at M orthogonally.*

Proof. Let A be the second point of intersection of $\mathfrak{C}, \mathfrak{C}_1$ (fig. 3). Let $\mathfrak{C}'$ be any chain tangent to $\mathfrak{C}$ at M. $\mathfrak{C}'$ will meet $\mathfrak{C}_1$ in a point A' distinct from M; for, if M were the only point common to $\mathfrak{C}'$, $\mathfrak{C}_1$, we should have two chains, $\mathfrak{C}$ and $\mathfrak{C}_1$, through A tangent to $\mathfrak{C}'$ at M (V, 3). The parabolic projectivity determined by $\pi(MMA) = MMA'$ leaves $\mathfrak{C}_1$ invariant and transforms $\mathfrak{C}$ into $\mathfrak{C}'$. Hence $\mathfrak{C}'$ and $\mathfrak{C}_1$ are orthogonal (Theorem 9); i.e., any chain tangent to $\mathfrak{C}$ at M meets $\mathfrak{C}_1$ orthogonally. That it meets any chain tangent to $\mathfrak{C}_1$ at M orthogonally follows by reasoning entirely similar to that just given.

COROLLARY. *Given a system S of chains mutually tangent at a point M, there exists a second system S_1 of chains mutually tangent at M such that every chain of S meets every chain of S_1 orthogonally.*

SECTION 5. **Analytic formulation.** The analytic criteria for the above classification of projectivities are now readily obtainable. If, in connection for example with the developments of A, Section 2 ff., we choose any three points as 01∞, the points of the line are made isomorphic with the system of ordinary complex numbers and ∞. The real numbers then correspond to the chain $|0\ 1\ \infty|$; the purely imaginary numbers to the chain $|0 i \infty|$, which is the chain through $0\ \infty$ orthogonal to $|0\ 1\ \infty|$.

Definition. The *absolute value* of a number x is the positive number in which the chain through x and about 0∞ meet $|01\infty|$. Two numbers then have the same absolute value, if they are on the same chain about 0∞.

Two numbers are said to have the same *amplitude*, if they are on the same chain through $0\,\infty$.

With this analytic representation the projectivities are given by the linear fractional transformations

$$\pi(x) = x' = \frac{ax + b}{cx + d}\,, \quad ad - bc \neq 0,$$

where x, x', a, b, c, d are complex numbers. If a projectivity, π, has two distinct double points x_1, x_2, and if x, x' is any pair of homologous points of π, it is a well-known theorem of projective geometry, that the double ratio

$$(1) \qquad \frac{x - x_2}{x - x_1} \cdot \frac{x' - x_1}{x' - x_2} = k$$

is constant for all pairs $x\,x'$. The projectivity π may therefore in this case be written

$$(2) \qquad \frac{x' - x_1}{x' - x_2} = k \cdot \frac{x - x_1}{x - x_2}.$$

The desired criteria are then obtained as follows:

π is parabolic, if we have $(d - a)^2 + 4bc = 0$.

The double points of π will be distinct, if we have $(d - a)^2 + 4bc \neq 0$.

If in (1) we regard x_1, x_2, x as fixed and x', k as variable, this defines a projectivity from x' to k, whereby x_1, x_2, x are transformed respectively into $0\,\infty\,1$. By this transformation the chain through x and $x_1\,x_2$ is transformed into $|0\,1\,\infty|$; whereas the chain through x and about x_1x_2 is transformed into the chain through 1 and about $0\,\infty$. It follows, then, that $x\,x'$ are on the same chain through x_1x_2, if and only if k is real: and that $x\,x'$ are on the same chain about $x_1\,x_2$, if and only if the absolute value of k is 1. This gives the result desired:

A projectivity (2) *is hyperbolic, if k is real;* elliptic, if the absolute value of k is* 1; *loxodromic, for any other values of k. An involution is characterized by the value $k = -1$*, which follows directly from the fact that an involution is both elliptic and hyperbolic, and the only real number different from 1 ($k = 1$ gives the identical projectivity) whose absolute value is 1 is -1; or from (1) and III, 4.

Since the necessary and sufficient condition that a projectivity $x' = (ax + b)/(cx + d)$ have real coefficients only is that it leave the chain $|0\ 1\ \infty|$ invariant, it follows readily that a projectivity can be transformed into one with real coefficients, if and only if it leaves a chain invariant. For, any projectivity transforming an invariant chain into $|0\ 1\ \infty|$ will transform the given projectivity into one leaving $|0\ 1\ \infty|$ invariant. This result may be stated as follows:

Every hyperbolic, elliptic, or parabolic projectivity can be represented analytically by a linear fractional transformation with real coefficients; and conversely, a linear fractional transformation with real coefficients is either hyperbolic, elliptic, or parabolic.

Section 6. **Groups of projectivities with invariant chain.**

In view of the simplicity of the reasoning it seems worth while to apply the results obtained above to the enumeration of the distinct types of continuous groups of projectivities which leave a chain invariant; or in other words, which may be represented analytically with real coefficients.

Lie's enumeration makes no distinction between real and complex: his parameters mean complex parameters. From this point of view there are four distinct types of continuous groups of projectivities, viz:

G_3: The set of all projectivities on a line.

G_2: The set of all projectivities leaving a given point invariant.

G_1: The set of all projectivities leaving two distinct points invariant.

G_1': The set of all parabolic projectivities leaving a given point invariant.

Every continuous group of projectivities in which the parameters are

* This definition of hyperbolic projectivities agrees e. g., with the definition of Enriques, *Vorlesungen über projective Geometrie* (Leipzig, 1903), p. 101; Goursat, *Cours d'analyse*, vol. 2 (Paris, 1905), p. 73; Burkhardt, *Theorie der analytischen Funktionen* (Leipzig, 1897), p. 38; with many authors a projectivity is hyperbolic only when k is real and *positive*: Osgood, *Funktionentheorie* (Leipzig, 1907), p. 225, and apparently the majority of books on the theory of functions.

complex can be transformed into one and only one of these types.* In seeking the different types of groups leaving a given chain $\mathfrak{C}$ invariant, we naturally regard two groups as belonging to the same type, if and only if one can be transformed into the other by a projectivity *leaving* $\mathfrak{C}$ *invariant*. Our previous results enable us to write down at once the following five distinct types:

RG_3 : The set of all (hyperbolic, elliptic, and parabolic) projectivities leaving $\mathfrak{C}$ invariant.

RG_2 : The set of all (hyperbolic and parabolic) projectivities leaving $\mathfrak{C}$ and a given point of $\mathfrak{C}$ invariant.

RGh_1 : The set of all (hyperbolic) projectivities leaving $\mathfrak{C}$ and two points of $\mathfrak{C}$ invariant.

RGe_1 : The set of all (elliptic) projectivities leaving $\mathfrak{C}$ and two points conjugate with respect to $\mathfrak{C}$ invariant.

RGp_1 : The set of all parabolic projectivities leaving $\mathfrak{C}$ and a given point of $\mathfrak{C}$ invariant.

That there can be no other types, follows readily from Lie's theorem † that if in a continuous group with real parameters the parameters are allowed to take on all complex values, there results a continuous group with complex parameters provided the transformations of the group are expressed by analytic functions.

Attention may also be called to the evident possibility of deriving synthetically much of the geometric theory of *discontinuous* groups of linear fractional substitutions on a complex variable ‡ by the methods exhibited in this paper. While probably no great simplification would arise from this method of treatment in the case of one variable, it is a significant fact that the corresponding analysis for linear fractional substitutions on two complex variables is very involved, while the geometry of two dimensional chains in a plane (which forms the subject of a paper to be published in the near future) is comparatively very simple. It seems likely therefore that the synthetic approach to the geometric theory of the groups of the automorphic functions

---

* Lie, *Vorlesungen über continuierliche Gruppen* (Leipzig, 1893), pp. 115 ff.

† Lie, *Theorie der Transformationsgruppen*, vol. 3, Leipzig, 1893, p. 362.

‡ Cf., for example, the classical paper by Poincaré, Sur les groupes fuchsiens, *Acta Mathematica*, vol. 1, (1881), p. 1.

of two independent variables will yield results of value. For any such approach the methods and results of the present paper will form a starting point. Finally, we would note the connection of the present paper with the inversion geometry in the plane. A further synthetic study of the transformation which consists in replacing each point by its inverse with respect to a chain gives a natural approach to the geometry of inversion and may well form the starting point of an investigation looking toward the setting up of a set of assumptions characterizing this geometry. The solution of this problem is greatly to be desired.*

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

*Cf. Kasner, The Present Problems of Geometry, *Bulletin of the American Mathematical Society*, vol. 11 (1905), p. 290.

NECESSARY AND SUFFICIENT CONDITIONS THAT AN ORDINARY DIFFERENTIAL EQUATION SHALL ADMIT A CONFORMAL GROUP

By L. I. Hewes

1. Introduction. We are to study in this paper a necessary and sufficient condition that an ordinary differential equation of the first order

$$\frac{dy}{dx} = a(x, y) \tag{1}$$

shall admit a continuous one-parameter conformal group. We shall assume the general properties of conformal transformation groups and in particular the following theorem due to Professor Bouton: *

A one-parameter group of conformal transformations with given path curves exists when and only when the given curves form an isothermal family.

We also assume the general methods laid down by Lie † and used by Tresse ‡ in the study of differential invariants.

It will be shown that under the infinite group of all conformal transformations there exist differential invariants formed from $a(x, y)$ and its partial derivatives. There are two of the third order and none of lower order. An infinite number of higher order are readily obtained from these two of lowest order. The desired condition for admission to occur is found to be dependent upon the nature of the invariants of the third and fourth orders.

2. The Differential Invariants. The discussion is greatly simplified by the introduction of circular coordinates §

$$u = x + iy, \qquad v = x - iy, \qquad\qquad i = \sqrt{-1}.$$

* *Bull. Am. Math. Soc.* vol. 11, p. 369.

† *Math. Ann.*, vol. 24, pp. 550-576.

‡ *Acta Math.* vol. 18, p. 76, also *Preisshrift, Jab. Gesell. zu Leipzig*, 1893.

§ F. Franklin, *Am. Journ. Math.*, vol. 12, 1890, p. 161.

The finite transformations of the infinite conformal group are then

$$u' = \Phi(u), \qquad v' = \Psi(v),$$

and the corresponding infinitesimal symbol may be written:

$$Uf: \qquad \phi(u)\frac{\partial f}{\partial u} + \psi(v)\frac{\partial f}{\partial v},$$

where $\phi(u)$ and $\psi(v)$ denote the first derivatives of $\Phi(u)$ and $\Psi(v)$ respectively. If we extend Uf to operate upon $\frac{dv}{du}$ we have

$$U'f: \qquad \phi(u)\frac{\partial f}{\partial u} + \psi(v)\frac{\partial f}{\partial v} + \pi(\psi'(v) - \phi'(u))\frac{\partial f}{\partial \pi},$$

when π denotes $\frac{dv}{du}$. In the new coordinates the differential equation (1) becomes

$$\frac{dv}{du} = m(u, v) \tag{2}$$

or more compactly

$$\pi - m = 0, \tag{2'}$$

and it preserves this form under conformal transformations. If we regard m as a variable transformed by the infinitesimal transformation $U'f$ its increment is determined by the identity

$$m(\psi' - \phi') - \delta m = 0.$$

Lie* has given formulas for computing the increments which the partial derivatives of m will also receive under $U'f$ and these become in the present case

$$\frac{\delta m_u}{\delta t} = \frac{\partial}{\partial u}\left(\frac{\delta m}{\delta t}\right) - m_u\phi', \qquad \frac{\delta m_v}{\delta t} = \frac{\partial}{\partial v}\left(\frac{\delta m}{\delta t}\right) - m_v\psi', \tag{I}$$

when the subscripts denote partial derivatives.

* l. c. Also Lie-Engel, vol. 1, p. 545.

By repeated use of these formulas we may compute the following table of increments (δt is omitted) of the successive partial derivatives of m up to the third order.

$$
\begin{aligned}
\delta m &= m(\psi' - \phi') \\
\delta m_u &= m_u(\psi' - 2\phi') - m\phi'' \\
\delta m_v &= - m_v\phi' + m\psi'' \\
\delta m_{uu} &= m_{uu}(\psi' - 3\phi') - 3m_u\phi'' - m\phi''' \\
\delta m_{uv} &= - 2m_{uv}\phi' + m_u\psi'' - m_v\phi'' \\
\delta m_{vv} &= - m_{vv}(\phi' + \psi') + m_v\psi'' + m\psi''' \\
\delta m_{uvv} &= - m_{uvv}(2\phi' + \psi') + m_{uv}\psi'' - m_{vv}\phi'' + m_u\psi''' \\
\delta m_{uuu} &= m_{uuu}(\psi' - 4\phi') - m_{uu}6\phi'' - 4m_u\phi''' - m\phi'''' \\
\delta m_{uuv} &= - 3m_{uuv}\phi' - 3m_{uv}\phi'' + m_{uu}\psi'' - m_v\phi''' \\
\delta m_{vvv} &= - m_{vvv}(\phi' + 2\psi') + 2m_v\psi''' + m\psi''''.
\end{aligned}
$$

Then the corresponding infinitesimal operator upon any function F of m and its partial derivatives above is therefore:

$$UF \equiv \delta_m \frac{\partial F}{\partial m} + \delta m_u \frac{\partial F}{\partial m_u} + \ldots + \delta m_{vvv} \frac{\partial F}{\partial m_{vvv}}.$$

We seek now those functions F of m and its partial derivatives which satisfy the condition:

$$(3) \qquad UF = 0.$$

Since this condition is to hold true for all conformal transformations, that is to say, for every ϕ and ψ, we find that the coefficients of ϕ and ψ and their derivatives in equation (3), must vanish. This gives rise to the following complete system of linear partial differential equations:

$$[1]\quad m\frac{\partial F}{\partial m} + 2m_u\frac{\partial F}{\partial m_u} + m_v\frac{\partial F}{\partial m_v} + 3m_{uu}\frac{\partial F}{\partial m_{uu}} + 2m_{uv}\frac{\partial F}{\partial m_{uv}} + m_{vv}\frac{\partial F}{\partial m_v^{v}}$$

$$+ 4m_{uuu}\frac{\partial F}{\partial m_{uuu}} + 3m_{uuv}\frac{\partial F}{\partial m_{uuv}} + 2m_{uvv}\frac{\partial F}{\partial m_{uvv}} + m_{vvv}\frac{\partial F}{\partial m_{vvv}} = 0$$

$$[2]\quad m\frac{\partial F}{\partial m_u} + 3m_u\frac{\partial F}{\partial m_{uu}} + m_v\frac{\partial F}{\partial m_{uv}} + 6m_{uu}\frac{\partial F}{\partial m_{uuu}} + 3m_{uv}\frac{\partial F}{\partial m_{uuv}}$$

$$+ m_{vv}\frac{\partial F}{\partial m_{uvv}} = 0$$

$$[3]\quad m\frac{\partial F}{\partial m_{uu}} + 4m_u\frac{\partial F}{\partial m_{uuu}} + m_v\frac{\partial F}{\partial m_{uuv}} = 0$$

$$[4]\quad m\frac{\partial F}{\partial m_{uuu}} = 0$$

$$[5]\quad m\frac{\partial F}{\partial m} + m_u\frac{\partial F}{\partial m_u} + m_{uu}\frac{\partial F}{\partial m_{uu}} - m_{vv}\frac{\partial F}{\partial m_{vv}} + m_{uuu}\frac{\partial F}{\partial m_{uuu}}$$

$$- m_{uvv}\frac{\partial F}{\partial m_{uvv}} - 2m_{vvv}\frac{\partial F}{\partial m_{vvv}} = 0$$

$$[6]\quad m\frac{\partial F}{\partial m_v} + m_u\frac{\partial F}{\partial m_{uv}} + m_v\frac{\partial F}{\partial m_{vv}} + m_{uu}\frac{\partial F}{\partial m_{uuv}} + m_{uv}\frac{\partial F}{\partial m_{uvv}} = 0$$

$$[7]\quad m\frac{\partial F}{\partial m_{vv}} + m_u\frac{\partial F}{\partial m_{uvv}} + 2m_v\frac{\partial F}{\partial m_{vvv}} = 0$$

$$[8]\quad m\frac{\partial F}{\partial m_{vvv}} = 0$$

We thus find that *until we use the derivatives of the third order*, there are not enough variables to give us any solutions. There are plainly two solutions of the system [1] $\cdots$ [8]. To obtain these we adopt as the most practical method the use of the solutions of one equation as new variables which are introduced into each of the remaining equations. This process repeated gives us finally one equation in three variables whose solutions are:

$$X_1 = \frac{1}{(m_u m_v - m m_{uv})^{\frac{1}{2}}}\left[\frac{m_u}{m^{\frac{1}{2}}} + \frac{1}{3}\,\frac{m^{\frac{1}{2}}(m m_{uuv} - m_v m_{uu})}{(m_u m_v - m m_{uv})}\right],$$

$$Y_1 = \frac{m^{\frac{1}{2}}}{(m_u m_v - m m_{uv})^{\frac{1}{2}}}\left[m_v + \frac{m(m m_{uvv} - m_u m_{vv})}{m_u m_v - m m_{uv}}\right].$$

These invariants may be given a simpler form as follows:

Let
$$a = \frac{\partial^2}{\partial u \partial v} \log m = \frac{m_{uv}\, m - m_u\, m_v}{m^2}.$$

The function a vanishes identically when and only when the integral curves form an isothermal family. This case is of no interest from the present view-point.

Write
$$X = -3iX_1 \quad \text{and} \quad Y = -iY_1,$$
then
$$X = \frac{1}{(ma)^{\frac{1}{2}}} \frac{\partial}{\partial u}\left(\log \frac{a}{m}\right), \quad Y = \left(\frac{m}{a}\right)^{\frac{1}{2}} \frac{\partial}{\partial v}\left(\log am\right).$$

The functions X and Y cannot both vanish identically unless $a \equiv 0$.

Now it is evident that we could calculate the increments of the derivatives of the next higher order viz:

$$\delta m_{uuuu},\ \delta m_{uuuv},\ \delta m_{uuvv},\ \delta m_{uvvv},\ \delta m_{vvvv},$$

and annex the corresponding additional equations to our above system and again seek the solutions. There would be ten equations with fifteen variables and hence three new solutions of the next higher order than X and Y. In general, if the number of invariants of the Kth order is N there will be $N + K$ of the $(K + 1)$st order. We do not, however, need to carry out such a process for we may obtain all the invariants required by a simpler method; namely, that of the "Differential Parameter."

We calculate the increments of the partial derivatives of either invariant, say of

$$X_u \text{ and } X_v\,*$$

under the infinitesimal transformation $U'f$. We find these increments by formulas I:

$$\frac{\delta X_u}{\delta t} = \frac{\partial}{\partial u} \frac{\delta X}{\delta t} - X_u \phi',$$

$$\frac{\delta X_v}{\delta t} = \frac{\partial}{\partial v} \frac{\delta X}{\delta t} - X_v \psi';$$

since X is an invariant function, $\delta X = 0$, and there results:

* It is assumed that neither X_u nor X_v vanishes identically.

$$\frac{\delta X_u}{\delta t} = - X_u\phi', \qquad \frac{\delta X_v}{\delta t} = - X_v\psi'.$$

We can now determine if there are invariants $\mathfrak{F}$ which involve

$$m, X_u, X_v,$$

for they must satisfy the relation

$$- X_u\phi' \frac{\partial \mathfrak{F}}{\partial X_u} - X_v\psi' \frac{\partial \mathfrak{F}}{\partial X_v} + m(\psi' - \phi') \frac{\partial \mathfrak{F}}{\partial m} = 0$$

for every ϕ' and ψ'. We have therefore to solve the linear partial differential equations

[9] $$X_u \frac{\partial \mathfrak{F}}{\partial X_u} + m \frac{\partial \mathfrak{F}}{\partial m} = 0,$$

[10] $$X_v \frac{\partial \mathfrak{F}}{\partial X_v} - m \frac{\partial \mathfrak{F}}{\partial m} = 0.$$

These equations form a complete system. Assuming that neither X_u nor X_v vanishes identically, a solution is

$$\frac{mX_v}{X_u}.^{*}$$

Similarly, using the invariant Y we should obtain the invariant

$$\frac{mY_v}{Y_u}.^{*}$$

We shall call these the first derived invariants and designate them as I and J respectively.

3. The Necessary Condition for Admission. Consider now the case that the differential equation

(2′) $$\pi - m = 0$$

admits a one-parameter conformal group whose infinitesimal transformation is

$$U'f: \qquad \phi \frac{\partial f}{\partial u} + \psi \frac{\partial f}{\partial v} + \pi(\psi' - \phi') \frac{\partial f}{\partial \pi}.$$

* The geometric interpretation of these invariants is immediately found to be the tangent of the angle between the integral curves and the curves $X = C$ or $Y = C$ respectively.

By the definition of admission the transformed differential equation is,

$$\pi' - m(u', v') = 0.$$

The general differential invariants X and Y are now invariant functions of u and v as are likewise I and J; they are all therefore as functions of u and v, functions of the unique integral invariant of the group and any one, set equal to a constant, must determine the path curves. They are therefore all functions of the same harmonic function,

$$U + V,$$

where U and V are functions of u and v alone, respectively.

Evidently *all* the differential invariants of higher orders submit to the same condition. We may then state as a necessary condition, which will be shown to be sufficient, the following:

A necessary condition that a first order ordinary differential equation shall admit a one-parameter conformal group is that the invariant X (or Y) and its first derived invariant I (or J) shall be functions of the same harmonic function.

4. The Sufficient Condition. Then $X = c$ determines the unique conformal group:

$$\overline{U'}f: \quad \rho \frac{\partial f}{\partial u} - \rho \frac{X_u}{X_v} \frac{\partial f}{\partial v} - \pi \left(\frac{\partial}{\partial v} \rho \frac{X_u}{X_v} + \frac{\partial}{\partial u} \rho \right) \frac{\partial f}{\partial \pi},$$

where ρ is a function of u and v satisfying the conditions

$$[4] \qquad \frac{\partial \rho}{\partial v} = 0, \qquad \frac{\partial}{\partial u} \rho \frac{X_u}{X_v} = 0.$$

From the second of these equations we have also

$$[4'] \qquad \frac{\partial \log \rho}{\partial u} = - \frac{\partial}{\partial u} \log \frac{X_u}{X_v}.$$

The necessary and sufficient condition that

$$(2') \qquad \pi - m = 0$$

shall admit $\overline{U'}f$ may be written

$$\rho \frac{m_u}{m} - \rho \frac{X_u}{X_v} \frac{m_v}{m} + \rho \frac{X_v X_{uv} - X_u X_{vv}}{X_v^2} + \rho_u = 0;$$

using the relation [4′] we have

$$[5] \qquad \frac{m_u}{m} - \frac{X_u}{X_v}\frac{m_v}{m} + \frac{X_v X_{uv} - X_u X_{vv}}{X_v^2} + \frac{X_u X_{uv} - X_v X_{uu}}{X_u X_v} = 0.$$

It is now easy to show that this condition [5] is satisfied whenever the assumed necessary condition of §3 is satisfied.

Multiply [5] by X_v and rearrange the terms, and there results

$$[5'] \qquad X_v\left(\frac{m_u}{m} + \frac{X_{uv}}{X_v} - \frac{X_{uu}}{X_u}\right) - X_u\left(\frac{m_v}{m} + \frac{X_{vv}}{X_v} - \frac{X_{uv}}{X_u}\right) = 0.$$

Equation [5′] may be given the determinant form:

$$[5''] \qquad \begin{vmatrix} \frac{\partial}{\partial u}\left[\log\left(m\frac{X_v}{X_u}\right)\right] & \frac{\partial}{\partial v}\left[\log\left(m\frac{X_v}{X_u}\right)\right] \\ \frac{\partial}{\partial u}X & \frac{\partial}{\partial v}X \end{vmatrix} = 0.$$

But [5″] is always satisfied when X and I are functions of the same harmonic function $U + V$. We have therefore the

THEOREM. *A necessary and sufficient condition that an ordinary differential equation of the first order shall admit a one-parameter conformal group is that a conformal invariant of the third order and its first derived invariant of the fourth order shall be functions of the same harmonic function.*

5. Interpretation of the Invariants. We may write the invariants

$$X = \frac{1}{(ma)^{\frac{1}{2}}}\frac{\partial}{\partial u}\log\frac{a}{m}, \qquad Y = \left(\frac{m}{a}\right)\frac{\partial}{\partial v}\log am$$

in the original coordinates x and y. We shall for that purpose adopt the notation

$$\tau = \text{arc tan } a,$$

$$\Delta\tau = \frac{\partial^2\tau}{\partial x^2} + \frac{\partial^2\tau}{\partial y^2},$$

and denote partial differentiation where convenient by subscripts.

We furthermore find it convenient to multiply X and Y by $(-2i)^{\frac{1}{2}}$.

We have then:

$$X' = \frac{1}{[\Delta\tau(1+a^2)]^{\frac{1}{4}}}\left[\frac{\Delta\tau_x + a\Delta\tau_y}{\Delta\tau} + \frac{2(a_y - aa_x)}{1+a^2}\right.$$
$$\left. + i\left(\frac{a\Delta\tau_x - \Delta\tau_y}{\Delta\tau} + \frac{2(aa_y + a_x)}{1+a^2}\right)\right],$$

$$Y' = \frac{1}{[\Delta\tau(1+a^2)]^{\frac{1}{4}}}\left[\frac{\Delta\tau_x + a\Delta\tau_y}{\Delta\tau} + \frac{2(a_y - aa_x)}{1+a^2}\right.$$
$$\left. - i\left(\frac{a\Delta\tau_x - \Delta\tau_y}{\Delta\tau} + \frac{2(aa_y + a_x)}{1+a^2}\right)\right].$$

Adding X' and Y' and subtracting, in order to obtain real forms, we have finally the two independent invariants of the third order:

$$\bar{X} = \frac{1}{[\Delta\tau(1+a^2)^{\frac{1}{4}}]}\left[\frac{a\Delta\tau_y + \Delta\tau_x}{\Delta\tau} - \frac{2(a_y - aa_x)}{1+a^2}\right],$$
$$\bar{Y} = \frac{1}{[\Delta\tau(1+a^2)]^{\frac{1}{4}}}\left[\frac{a\Delta\tau_x - \Delta\tau_y}{\Delta\tau} + \frac{2(aa_y + a_x)}{1+a^2}\right].$$

Denoting the curvature of the integral curves of the original differential equation by κ, we have:

$$\kappa = \frac{a_x + aa_y}{[1+a^2]^{\frac{3}{2}}}.$$

Similarly writing the curvature of the orthogonal trajectories,

$$\bar{\kappa} = \frac{aa_x - a_y}{(1+a^2)^{\frac{3}{2}}},$$

we may then write $\Delta\tau = \frac{d\kappa}{ds} + \frac{d\bar{\kappa}}{d\bar{s}}$ where s and $\bar{s}$ denote arcs on the integral curve and the orthogonal trajectory respectively. With this understanding we may write the invariants $\bar{X}$ and $\bar{Y}$ as functions of the curvatures κ and $\bar{\kappa}$ and their first and second derivatives with respect to the arcs.

$$\bar{X} = \frac{1}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{4}}}\left[\frac{\kappa_{ss} + \bar{\kappa}_{s\bar{s}}}{\kappa_s + \bar{\kappa}_{\bar{s}}} - \frac{2\bar{\kappa}}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{4}}}\right],$$
$$\bar{Y} = \frac{1}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{4}}}\left[\frac{\kappa_{s\bar{s}} + \bar{\kappa}_{\bar{s}\bar{s}}}{\kappa_s + \bar{\kappa}_{\bar{s}}} + \frac{2\kappa}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{4}}}\right].$$

The fourth order invariants may be given similar form. In terms of the original variables x, y, a, a differential parameter is

$$\frac{a\Omega_y + \Omega_x}{\Omega_y - a\Omega_x},$$

where Ω is *any* invariant. Hence we may write as the two fourth order invariants,

$$\bar{I} = \frac{a\bar{X}_y + \bar{X}_x}{X_y - aX_x}, \qquad J = \frac{aY_y + Y_x}{Y_x + aY_y}.$$

Expressing these invariants in terms of the curvatures κ and $\bar{\kappa}$ and their derivatives we have,

$$\bar{I} = P/Q,\quad J = R/S,$$

where

$$P = 2\left[\frac{\partial^2}{\partial s^2}\log(\kappa_s + \bar{\kappa}_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \kappa_s)\right\}^2 + \frac{2\kappa(\kappa_{ss} + \kappa_{ss})}{(\kappa_s + \kappa_{\bar{s}})^{\frac{3}{2}}} - \frac{2\kappa_s}{(\kappa_s + \kappa_{\bar{s}})^{\frac{1}{2}}}\right]$$

$$+\left(\frac{1 - a^2}{a}\right)\left[\frac{\partial^2}{\partial s \partial s}\log(\kappa_s + \bar{\kappa}_s) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \kappa_s)\frac{\partial}{\partial s}\log(\kappa_s + \kappa_s)\right\}\right.$$

$$\left. + \frac{2\kappa(\kappa_{s\bar{s}} + \kappa_{ss})}{(\kappa_s + \kappa_s)^{\frac{3}{2}}} - \frac{2\kappa_{\bar{s}}}{(\kappa_s + \kappa_{\bar{s}})^{\frac{1}{2}}}\right]$$

$$Q = \left(\frac{1 - a^2}{a}\right)\left[\frac{\partial^2}{\partial s^2}\log(\kappa_s + \kappa_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \kappa_{\bar{s}})\right\}^2 + \frac{2\kappa(\kappa_{ss} + \bar{\kappa}_{\bar{s}s})}{(\kappa_s + \kappa_s)^{\frac{3}{2}}}\right.$$

$$\left. - \frac{2\bar{\kappa}_s}{(\kappa_s + \kappa_{\bar{s}})^{\frac{1}{2}}}\right] - 2\left[\frac{\partial^2}{\partial s \partial s}\log(\kappa_s + \bar{\kappa}_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \bar{\kappa}_{\bar{s}})\frac{\partial}{\partial s}\log(\kappa_s + \bar{\kappa}_s)\right\}\right.$$

$$\left. + \frac{2\kappa(\kappa_{s\bar{s}} + \kappa_{ss})}{(\kappa_s + \kappa_s)^{\frac{3}{2}}} - \frac{2\kappa_{\bar{s}}}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{2}}}\right],$$

$$R = 2\left[\frac{\partial^2}{\partial s \partial s}\log(\kappa_s + \bar{\kappa}_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \kappa_s)\frac{\partial}{\partial s}\log(\kappa_s + \kappa_s)\right\} - \frac{2\kappa(\kappa_{ss} + \kappa_{ss})}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{3}{2}}}\right.$$

$$\left. - \frac{2\kappa_s}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{2}}}\right] + \left(\frac{1 - a^2}{a}\right)\left[\frac{\partial^2}{\partial s^2}\log(\kappa_s + \kappa_s) - \frac{1}{2}\left\{\frac{\partial}{\partial s}\log(\kappa_s + \kappa_{\bar{s}})\right\}^2\right.$$

$$\left. - \frac{2\kappa(\kappa_{s\bar{s}} + \kappa_{\bar{s}\bar{s}})}{(\kappa_s + \kappa_s)^{\frac{3}{2}}} + \frac{2\kappa_{\bar{s}}}{(\kappa_s + \kappa_{\bar{s}})^{\frac{1}{2}}}\right],$$

$$S = \left(\frac{1-a^2}{a}\right)\left[\frac{\partial^2}{\partial s \partial \bar{s}} \log(\kappa_s + \bar{\kappa}_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial \bar{s}} \log(\kappa_s + \bar{\kappa}_{\bar{s}}) \frac{\partial}{\partial s} \log(\kappa_s + \bar{\kappa}_{\bar{s}})\right\}\right.$$
$$\left. - \frac{2\kappa(\kappa_{ss} + \bar{\kappa}_{\bar{s}\bar{s}})}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{3}{2}}} + \frac{2\kappa_s}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{2}}}\right] - 2\left[\frac{\partial^2}{\partial s^2} \log(\kappa_s + \bar{\kappa}_{\bar{s}}) - \frac{1}{2}\left\{\frac{\partial}{\partial \bar{s}} \log(\kappa_s + \bar{\kappa}_{\bar{s}})\right\}^2\right.$$
$$\left. - \frac{2\kappa(\kappa_{ss} + \bar{\kappa}_{\bar{s}\bar{s}})}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{3}{2}}} + \frac{2\kappa_s}{(\kappa_s + \bar{\kappa}_{\bar{s}})^{\frac{1}{2}}}\right].$$

6. Conclusion. It is possible to write down the form that the function a must take in the differential equation

$$(1) \qquad \frac{dy}{dx} = a(x, y)$$

in order that the equation shall admit a conformal group. Since the path curves $\omega = c$, of such a group must individually cut each integral curve under a constant angle along that curve we have at once

$$[6] \qquad \frac{\dfrac{dy}{dx} + \dfrac{\omega_x}{\omega_y}}{1 - \dfrac{dy}{dx}\dfrac{\omega_x}{\omega_y}} = F(\omega),$$

where ω is a solution of Laplace's equation,

$$\omega_{xx} + \omega_{yy} = 0.$$

Solving [6] for $\dfrac{dy}{dx}$, we obtain

$$[7] \qquad \frac{dy}{dx} = \frac{-\omega_x + F(\omega)\omega_y}{\omega_y + F(\omega)\omega_x}.$$

The corresponding infinitesimal conformal transformation is

$$Uf: \quad \frac{\omega_y}{\omega_x^2 + \omega_y^2}\frac{\partial f}{\partial x} - \frac{\omega_x}{\omega_x^2 + \omega_y^2}\frac{\partial f}{\partial y}.$$

It will be found that whenever a has the form in equation [7] that X, Y, I, and J are functions of ω, or in other words the form [7] is necessary and sufficient.

YALE UNIVERSITY,
JULY, 1908.

THE 3-SPACE $PG(3, 2)$ AND ITS GROUP.*

BY GEORGE M. CONWELL

Introduction. An n-space in which the points are determined by homogeneous coordinates will contain only a finite number of points if the coordinates are restricted to be integers reduced to a modulus. In this paper the modulus 2 is used, i. e., the only numbers possible are 0 and 1.

The configuration of the "real" points, i. e., those whose coordinates are expressed in terms of 0 or 1, is studied in detail.

The complete projective group of collineations and dualities of the 3-space is shown to be of order $|\underline{8}$ and to have as a sub-group the linear homogeneous group L. H. G. $\overset{15}{\underset{2}{|\underline{8}}}$, this sub-group consisting of the projective collineations of the 3-space. The line coordinates for the lines of the 3-space are introduced and the linear complexes and congruences obtained. These line coordinates can also be considered as the coordinates of points in a 5-space. To every transformation of the 3-space there corresponds a transformation of the 5-space. In the 5-space, there are determined 8 sets of 7 points each, "heptads," by means of which is established the isomorphism of the linear homogeneous group L. H. G. $\overset{15}{\underset{2}{|\underline{8}}}$ with the alternating group on 8 letters.

An 8 letter notation is derived from the "heptads" for the points, lines, and planes of the 3-space.

The geometry gives a complete solution of Kirkman's School Girl Problem and is related to several functions which are of importance in the Galois theory of equations. The configuration of the 3-space is the same as that studied by Moore in this connection.†

The Configuration of the "Real" Points. A point is determined in a 3-space by 4 homogeneous coordinates (a, b, c, d) and when the numbers

*Oswald Veblen and W. H. Bussey, Finite Projective Geometries. *Transactions of the American Mathematical Society*, vol. 7 (1906), pp. 241-259.

† E. H. Moore, Concerning the General Equations of the Seventh and the Eighth Degrees *Mathematische Annalen*, vol. 51 (1899), pp. 417-444.

a, b, c, d, are integers and reduced modulo 2 we obtain 15 points excluding the combination $(0, 0, 0, 0)$; these are called the "real" points of the 3-space.

If $P_{a_1} = (a_1, b_1, c_1, d_1)$ and $P_{a_2} = (a_2, b_2, c_2, d_2)$ are any two points, the points of the line joining them are given by

$$P_{\lambda a_1 + \eta a_2} = (\lambda a_1 + \eta a_2, \lambda b_1 + \eta b_2, \lambda c_1 + \eta c_2, \lambda d_1 + \eta d_2);$$

the only possible sets of values for (λ, η) are $(1, 0)$, $(0, 1)$, $(1, 1)$, hence the number of points on a line is 3. The points of a line may thus be denoted by P_{a_1}, P_{a_2}, $P_{a_1+a_2}$. Consider a fourth point $P_{a'}$, which is not on the line P_{a_1}, P_{a_2}, $P_{a_1+a_2}$; evidently $P_{a_1+a'}$, $P_{a_2+a'}$, $P_{a_1+a_2+a'}$, are points of the lines joining $P_{a'}$ to the points of the line P_{a_1}, P_{a_2}, $P_{a_1+a_2}$.

These 7 points are all the points of a plane, since any 2 points P_x and P_y determine a collinear point P_{x+y}, which is contained among the 7. The configuration of the 7 points of a plane is that of a complete quadrangle in which the diagonal points are collinear. The number of lines in a plane is 7 as there are 7 points and 3 lines through each point.

In the 3-space the number of "real" planes is the same as the number of "real" points, since for each possible set of point coordinates there is a possible set of plane coordinates. Consider the 7 points of a plane which can be denoted by

$$P_{a_1}, P_{a_2}, P_{a_1+a_2}, P_{a'}, P_{a_1+a'}, P_{a_2+a'}, P_{a_1+a_2+a'},$$

and a point Pa'' not upon the plane; this will determine 7 other points

$$P_{a_1+a''}, P_{a_2+a''}, P_{a_1+a_2+a''}, P_{a'+a''}, P_{a_1+a'+a''}, P_{a_2+a'+a''}, P_{a_1+a_2+a'+a''}.$$

These 15 points constitute all the "real" points of the 3-space, for any 2 points P_x and P_y determine a collinear point P_{x+y}, which is included among the 15 and any 3 non-collinear points P_x, P_y, P_z determine the 7 points of a plane, which can be denoted by

$$P_x, P_y, P_{x+y}, P_z, P_{x+z}, P_{y+z}, P_{x+y+z}.^*$$

Any 4 non-coplanar points determine 15 points. This process of "doubling" from a point without an n-space always leads to an $(n + 1)$-space or from a $PG(n, 2)$ to a $PG(n + 1, 2)$.

Through each point there are 7 lines, for take any point and a plane not

* E. H. Moore, loc. cit.

containing it, the 7 lines joining this point to the points of a plane contain all 15 points. The number of lines is 35, since there are 15 points and 7 lines through each point and 3 points on a line: $7.15/3 = 35$.

Of the 15 planes 3 pass through each line. Consider a plane and a point outside the plane, the point with the 7 lines of the plane determines 7 distinct planes, one of which contains any 2 points of the 3-space, hence there are 7 planes through each point.

The whole configuration can be exhibited in the table *

	S_0	S_1	S_2
S_0	15	7	7
S_1	3	35	3
S_2	7	7	15

S_0 is a point, S_1 a line, S_2 a plane and in general S_n is an n-space. A number n in the ith row and the ith column gives the number of i-spaces in the given space, while a number n in the ith column and the kth row gives the number of i-spaces which are united with a k-space.

By the use of the transformation T of period 15

$$T \quad \begin{aligned} x_1' &= x_1 +   + x_4 \\ x_2' &= x_1 + x_2 + x_4 \\ x_3' &= x_1 + x_2 + x_3 \\ x_4' &= x_1 + x_2 + x_3 + x_4 \end{aligned} \qquad \text{or} \qquad \begin{aligned} x_1 &= x_1' + x_3' + x_4' \\ x_2 &= x_1' + x_2' \\ x_3 &= x_2' + x_4' \\ x_4 &= x_3' + x_4' \end{aligned}$$

all the points can be obtained from a single one, say (1, 1, 1, 1), and all the planes from a single one, say $x_1 + x_2 + x_3 + x_4 = 0$.

1	(1,1,1,1)	I	$x_1 + x_2 + x_3 + x_4 = 0$
2	(0,1,1,0)	II	$x_4 = 0$
3	(0,1,0,0)	III	$x_3 + x_4 = 0$
. . .	. . .	. . .	. . . .

* E. H. Moore, Tactical Memoranda I–III. *American Journal of Mathematics*, vol. 18 (1896), pp. 264–303.

Hence denoting the points by the numbers 1 to 15, the point i is obtained from the point 1 (1,1,1,1) by the transformation T^i. All the planes are obtained from one expressed through its 7 points as 1 (1,2,5,7,12,13,14) by the transformation T (1,2,3,4, $\cdots$, 14,15) and its powers.

And all the lines can be obtained from 3, each expressed in terms of its 3 points, (1, 2, 5), (1, 3, 9), (1, 6, 11), by means of the same transformation T. The first two lines have each 15 conjugates while the third has 5 under the transformation T and its powers.

Introduction of Line Coordinates. Two points Pa and Pa' of the 3-space determine a line: P_a (a_1, a_2, a_3, a_4) and $P_a'(a_1', a_2', a_3', a_4')$.

There are 12 determinants p_{ik} of the form

$$p_{ik} = \begin{vmatrix} a_i & a_\kappa \\ a_i' & a_\kappa' \end{vmatrix}$$

which are related two by two, $p_{i\kappa} = p_{\kappa i}$ (since $p = -p$ modulo 2).

The 6 quantities $(p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34})$ are the Plücker line coordinates and completely determine a line. They are related by the condition

$$p_{12}p_{34} + p_{13}p_{24} + p_{14}p_{23} = 0.^*$$

The line coordinates of the 35 lines can be obtained from those of the 3 lines (1, 2, 5), (1, 3, 9), (1, 6, 11) by means of a transformation P on the p's which is induced by the transformation T:

$$\begin{array}{llllll}
p_{12}' = p_{12} & & & & + p_{24} & \\
p_{13}' = p_{12} & + p_{13} & + p_{14} & & + p_{24} & + p_{34} \\
p_{14}' = p_{12} & + p_{13} & & & + p_{24} & + p_{34} \\
p_{23}' = & + p_{13} & + p_{14} & + p_{23} & + p_{24} & + p_{34} \\
p_{24}' = & + p_{13} & & + p_{23} & & + p_{34} \\
p_{34}' = & & p_{14} & + & + p_{24} & + p_{34}
\end{array}$$

The 3 lines and their line coordinates are respectively (1, 2, 5), (1, 3, 9), (1, 6, 11), and (1,1,0,0,1,1), (1,0,0,1,1,0), (0,1,0,1,0,1).

It is evident that the line coordinates may be looked upon as points in a

*C. M. Jessop, *Treatise on the Line Complex*, page 17.

5-space, a $PG(5, 2)$. In such a space there are $2^6 - 1 = 63$ points excluding $(0,0,0,0,0,0)$. Of these 63 points, the 35 representing lines of the 3-space lie upon a surface, S,

$$p_{12}p_{34} + p_{13}p_{24} + p_{14}p_{23} = 0.$$

The lines whose coordinates satisfy a linear equation constitute a linear complex, which can be either degenerate or non-degenerate.

A degenerate complex consists of all the lines which meet a given line, called the axis, together with this axis. Given two lines p and p' with the line coordinates $(p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34})$ and $(p'_{12}, p'_{13}, p'_{14}, p'_{23}, p'_{24}, p'_{34})$, the condition that they may intersect is

$$p_{12}p'_{34} + p_{13}p'_{24} + p_{14}p'_{23} + p_{23}p'_{14} + p_{24}p'_{13} + p_{34}p'_{12} = 0.$$

Hence this condition, if the p's are considered constant, is the equation of a degenerate complex, consisting of all lines meeting p' and of p' itself. The equation is satisfied for p' since the equation then reduces to the condition S.

Any linear relation among the p's, as

$$ap_{12} + bp_{13} + cp_{14} + dp_{23} + ep_{24} + fp_{34} = 0,$$

can be looked upon as the polar of the point (f,e,d,c,b,a) with respect to the surface S and if this point in the 5-space represents a line in the 3-space, then the complex determined will be degenerate; if it does not represent a line, the complex is non-degenerate. Hence the polar of a point on the surface S, i. e., the tangent 4-space at the point to the surface S, has in common with the surface S the points representing the lines of a degenerate complex, while the polar of a point off the surface S determines in a similar way a non-degenerate complex. The number of degenerate complexes is equal to the number of points in the 5-space which represent lines in the 3-space, that is 35. They may all be obtained from the following three by the transformation P and its powers:

$$p_{12} + p_{13} + p_{24} + p_{34} = 0, \quad p_{12} + p_{23} + p_{34} = 0, \quad p_{13} + p_{23} + p_{34} = 0.$$

A non-degenerate complex is a system of lines such that through every point of the 3-space there is a flat pencil of lines and all the lines in every plane pass through a point. The proof is the same as in the ordinary case.*

* Jessop, loc. cit., page 25.

The lines whose coordinates satisfy two linear equations and hence are common to two complexes, constitute a congruence. Consider the two complexes C and C':

$$ap_{12} + bp_{13} + cp_{14} + dp_{23} + ep_{24} + fp_{34} = 0,$$
$$a'p_{12} + b'p_{13} + c'p_{14} + d'p_{23} + e'p_{24} + f'p_{34} = 0;$$

these determine a third complex, $C'' = C + C'$,

$$(a + a')\,p_{12} + (b + b')\,p_{13} + (c + c')\,p_{14} + (d + d')\,p_{23} + (e + e')\,p_{24} + (f + f')\,p_{34} = 0.$$

C, C' and C'' are degenerate if the points (a,b,c,d,e,f), (a',b',c',d',e',f') and $(a + a', b + b', c + c', d + d', e + e', f + f')$ respectively are upon the surface S. The condition that this last point is on the surface reduces to

$$af' + a'f + be' + b'e + cd' + c'd = 0.$$

This is the condition that the axes of C and C' intersect. Hence if the three complexes of a family are all degenerate, the congruence determined consists of 11 lines meeting the two intersecting axes of the complexes C and C'. If two of the complexes of the family are degenerate, while the third is non-degenerate, the congruence consists of 9 lines meeting the two non-intersecting axes of the degenerate complexes. If only one of the complexes of the family is degenerate, then the non-degenerate complexes contain the axis of the degenerate complex, the congruence consists of 7 lines, the axis and 6 lines meeting it, one line through each point of the 3-space. This is a degenerate congruence with an axis. When the three complexes of a family are all non-degenerate, there is a non-degenerate congruence determined, which consists of 5 lines such that there is one and but one through each point of the 3-space.

The Group of the 3-space. Every projective collineation* in the 3-space $PG(3, 2)$ is represented by a transformation T on the point coordinates

$$T:\quad x'_r = \sum_{i=1}^{i=4} a_{ri} x_i \qquad r = (1, 2, 3, 4).$$

* O. Veblen and W. H. Bussey, loc. cit., p. 253.

This transformation is completely determined if we know into what points the vertices of the tetrahedron of reference are transformed; (1,0,0,0) goes into $(a_{11}, a_{21}, a_{31}, a_{41})$, (0,1,0,0) goes into $(a_{12}, a_{22}, a_{32}, a_{42})$, etc.

In order that the transformation may have an inverse, the determinant of the transformation must be different from zero, and this is the condition that the four points $(a_{1i}, a_{2i}, a_{3i}, a_{4i})$, $(i = 1,2,3,4)$ into which the vertices of the tetrahedron are transformed shall not be coplanar. Hence the number of transformations of the group of projective collineations is equal to the number of ways four non-coplanar points can be chosen, i. e., $15 \cdot 14 \cdot 12 \cdot 8 = 20,160 = \lfloor 8/2$. The dualities or polarities of the 3-space are transformations of the form

$$x_r = \sum_{i=}^{i=4} a_{ri} u, \qquad r = (1, 2, 3, 4),$$

where the x's are point coordinates and the u's plane coordinates.

These transformations have the property of changing points into planes and vice versa, but change lines into lines. The number of these dualities is evidently the same as the number of projective collineations $\lfloor 8/2$. The order of the complete projective group, consisting of all projective collineations and dualities, is $\lfloor 8$.

Since a transformation of the complete projective group changes lines into lines, every transformation of this group determines a transformation on the line coordinates. Conversely every transformation on the line coordinates which changes lines into lines corresponds to a transformation of the complete projective group. Let the transformation on the p's be of the form

$$p'_r = \sum_{i=1}^{i=6} a_{ri} p_i, \qquad r = (1, 2, 3, 4, 5, 6),$$

The lines with the coordinates (1,0,0,0,0,0), (0,1,0,0,0,0), (0,0,1,0,0,0), (0,1,1,0,0,0), (1,0,1,0,0,0) and (1,1,0,0,0,0), all pass through the point (1,0,0,0) and must be transformed into lines, hence we must have

$$(A) \quad a_{1i}a_{6i} + a_{2i}a_{5i} + a_{3i}a^{4i} = 0, \quad (i = 1, 2, 3),$$

$$(B) \quad (a_{1i} + a_{1j})(a_{6i} + a_{6j}) + (a_{2i} + a_{2j})(a_{5i} + a_{5j}) + (a_{3i} + a_{3j})(a_{4i} + a_{4j}) = 0.$$

(B) reduces by means of (A) to (B'):

$$(B') \quad a_{1i}a_{6j} + a_{1j}a_{6i} + a_{2i}a_{5j} + a_{2j}a_{5i} + a_{3i}a_{4j} + a_{3j}a_{4i} = 0, \ (ij = 1, 2, 3).$$

(B') is the condition that the three lines (1,0,0,0,0,0), (0,1,0,0,0,0), (0,0,1,0,0,0), which are not in a plane, must be transformed into lines which meet two by two, hence they must pass through a point or lie in a plane. Hence all the lines through a point must be transformed into lines through a point, a projective collineation, or be transformed into the lines of a plane, a duality. The simplest transformation on the p's which is a duality is

$$p_1' = p_6, \quad p_2' = p_5, \quad p_3' = p_4, \quad p_4' = p_3, \quad p_5' = p_2, \quad p_6' = p_1.$$

The Configuration of the 28 Points not upon the Surface S in the 5-space. Interpreting the results of section 2 in terms of the 5-space we see that taking the polar of a point in the 5-space with respect to the surface S determines a non-degenerate complex for the 3-space. Taking the polars of three points of a line in the 5-space determines a non-degenerate congruence for the 3-space. Hence the number of non-degenerate congruences is the same as the number of lines in the 5-space, which are wholly off the surface S. Whatever is true for a single point of the 28 is true for all, as any point can be transformed into any other. The number of lines wholly off the surface S which can be drawn through the point (1,0,0,0,0,1) is 6 and the same number can be drawn through each of the 28 points, hence there are $28 \cdot 6/3 = 56$ lines wholly off the surface S. The number of non-degenerate congruences is thus 56.

Consider the 3 points of a line as $A = (1,0,0,0,0,1)$, $B = (1,0,1,1,0,0)$, $C = (0,0,1,1,0,1)$, which are wholly off the surface S. Through each of these points there are 5 other lines wholly off the surface S. These lines are denoted below in terms of their 3 points, the 3 points in a row are the 3 points of a line.

Lines through A	Lines through B	Lines through C
(1,0,0,0,0,1)(1,0,1,1,0,0)(0,0,1,1,0,1)	(1,0,0,0,0,1)(1,0,1,1,0,0)(0,0,1,1,0,1)	(1,0,0,0,0,1)(1,0,1,1,0,0)(0,0,1,1,0,1)
(1,0,0,0,0,1)(0,1,1,1,0,1)(1,1,1,1,0,0)	(0,1,1,1,0,1)(1,0,1,1,0,0)(1,1,0,0,0,1)	(1,1,1,1,0,0)(1,1,0,0,0,1)(0,0,1,1,0,1)
(1,0,0,0,0,1)(1,1,0,0,1,0)(0,1,0,0,1,1)	(0,1,0,0,1,1)(1,0,1,1,0,0)(1,1,1,1,1,1)	(1,1,1,1,1,1)(1,1,0,0,1,0)(0,0,1,1,0,1)
(1,0,0,0,0,1)(0,0,1,1,1,1)(1,0,1,1,1,0)	(0,1,1,1,1,1)(1,0,1,1,0,0)(1,0,0,0,1,1)	(1,0,0,0,1,1)(1,0,1,1,1,0)(0,0,1,1,0,1)
(1,0,0,0,0,1)(1,1,1,0,1,0)(0,1,1,0,1,1)	(0,1,0,1,1,0)(1,0,1,1,0,0)(1,1,1,0,1,0)	(0,1,0,1,1,0)(0,1,1,0,1,1)(0,0,1,1,0,1)
(1,0,0,0,0,1)(0,1,0,1,1,1)(1,1,0,1,1,0)	(1,1,0,1,1,0)(1,0,1,1,0,0)(0,1,1,0,1,0)	(0,1,1,0,1,0)(0,1,0,1,1,1)(0,0,1,1,0,1)

It will be observed that when two points as A and B of the line ABC are chosen, that each of the 5 other lines through A is met by a single line of the 5 other lines through B. The 5 points determined by the intersection of these pairs of lines together with the 2 points A and B we designate a "heptad." It will be found that the $7\cdot 6/2 = 21$ lines joining these 7 points by pairs are wholly off the surface S. If any 2 points of the heptad are chosen and these points are used in the same way as A and B for the determination of a heptad, the same heptad will be determined. Each of the pairs of points AB, BC, CA determines a heptad so that each point of the 28 points off the surface S belongs to 2 heptads. The number of heptads is thus $28\cdot 2/7 = 8$. No 6 of the points of a heptad are in the same 4-space, no 5 in the same 3-space, no 4 in the same 2-space, no 3 in the same 1-space.

If we take the generators of the group of projective collineations of the 3-space, the $L.H.G\left|\begin{smallmatrix}15\\8\\2\end{smallmatrix}\right.$, the matrices below are the matrices of the 6 generators of the group.*

E_1	E_2	E_3	E_4	E_5	E_6
1110	0101	0110	1110	0101	0011
0111	0110	0011	0100	0011	0110
1111	0010	1011	0010	1111	0010
1011	1110	1111	0101	1011	1001

From these can be calculated the corresponding transformations in terms of the p's. Let the 8 heptads be given the numbers 1 to 8; each heptad is determined by two of its points.

(1,0,1,1,0,0) and (1,1,1,1,1,1) determine heptad 1
(0,0,1,1,0,0) and (1,1,1,0,1,0) determine heptad 2
(1,1,0,0,1,0) and (0,1,0,1,1,1) determine heptad 3
(0,0,1,1,1,1) and (0,1,1,1,0,0) determine heptad 4
(1,0,1,0,1,1) and (0,1,1,1,0,1) determine heptad 5
(1,0,0,0,0,1) and (1,0,1,1,0,0) determine heptad 6
(1,1,1,0,0,1) and (0,0,1,1,0,0) determine heptad 7
(0,0,1,1,1,0) and (1,1,0,0,1,0) determine heptad 8.

* E. H. Moore, loc. cit., page 435.

The transformations in terms of the p's applied to the 28 points off the surface S permute the heptads among themselves. To the 6 generators above correspond the permutations on the heptads,

$$E_1 = (4,7)(7,6),\ E_2 = (4,7)(6,8),\ E_3 = (4,7)(8,5),$$
$$E_4 = (4,7)(1,5),\ E_5 = (4,7)(1,2),\ E_6 = (4,7)(2,3);$$

and these are the generators of the alternating group on 8 letters.

Hence the $L.H.G \left| \overset{15}{\underset{2}{8}} \right.$ and the alternating group on 8 letters are isomorphic.*

A transformation on the p's which is a duality is equivalent to a transformation of the symmetric group on 8 letters, for example, the duality of page 67 is equivalent to the transformation $(2,7)(3,6)(4,5)$ on the heptads. To the fact that by the addition of one duality to the group of projective collineations the complete projective group can be generated, corresponds the fact that by the addition of one uneven transformation to the alternating group the symmetric group is generated. The complete projective group of the 3-space, $PG(3,2)$, is isomorphic with the symmetric group on 8 letters.

By means of the heptads, it is possible to assign an 8 letter notation to the points of the 5-space which are off the surface S, and to study their configuration. We have pointed out the fact that each point off the surface S belongs to 2 heptads. The notation for a point can thus be taken as (a,b), $a \neq b$, and there are $8 \cdot 7/2 = 28$ combinations of this form corresponding to the 28 points off the surface S. The notation for the points of a line off the surface S is $(ab)(bc)(ca)$, a, b, c distinct; for the points of a line off the surface S by pairs determine three heptads. Thus the two letter notations for the points must by pairs have a letter in common. There are thus $8 \cdot 7 \cdot 6/1 \cdot 2 \cdot 3 = 56$ lines off the surface S. The maximum number of points which a plane can have off the surface S is six. Such a plane in terms of its points has the notation $(ab)(bc)(ca)(ad)(bd)(cd)$. A line off the surface S has the notation $(ab)(bc)(ca)$. For any point off the surface S has the notation (xy) and in order that the line determined by this point with (ab) may have its third point off the surface S, x or y must equal a or b since (ab) and (xy) must belong to a common heptad. Let $x = a$ then $y = d$, $d \neq a \neq b \neq c$. Then the two other

*First proved by Jordan, *Traité des substitutions*, No. 516. E. H. Moore has given a proof based on the same system of generators, loc. cit., page 432.

points of the plane which are off the surface S are (bd) and (cd). There are $8\cdot 7\cdot 6\cdot 5/1\cdot 2\cdot 3\cdot 4 = 70$ planes of this kind.

In a similar manner the 3-spaces and the 4-spaces with the maximum number of points off the surface S can be determined. The configuration of the n-spaces with the maximum number of points off the surface S is given below in a table whose interpretation is the same as that of the table page 62

	S_0	S_1	S_2	S_3	S_4
S_0	28	6	15	20	15
S_1	3	56	5	10	10
S_2	6	4	70	4	6
S_3	10	10	5	56	3
S_4	15	20	15	6	28

The Configuration of the Thirty-Five Points upon the Surface S in the 5-space. The surface S has as its equation

$$p_{12}p_{34} + p_{13}p_{24} + p_{14}p_{23} = 0$$

and is satisfied by the 35 points of the 5-space representing lines in the 3-space. Consider any two points upon the surface S, $(a_1, b_1, c_1, d_1, e_1, f_1)$ and $(a_2, b_2, c_2, d_2, e_2, f_2)$ the third point of the line, $(a_1 + a_2, b_1 + b_2, c_1 + c_2, d_1 + d_2, e_1 + e_2, f_1 + f_2)$ will be upon the surface S if

$$(a_1 + a_2)(f_1 + f_2) + (b_1 + b_2)(e_1 + e_2) + (c_1 + c_2)(d_1 + d_2) = 0$$

or if
$$a_1f_2 + a_2f_1 + b_1e_2 + b_2e_1 + c_1d_2 + c_2d_1 = 0.$$

This latter is the condition that the two lines in the 3-space represented by the first two points intersect. Any two points of a line can be taken as the first two, hence the 3 lines in the 3-space which are represented by the points of a line in the 5-space must meet by pairs in the 3-space and thus lie in a plane, or pass through a point.

Let $A = (a_1, a_2, a_3, a_4)$, $B = (b_1, b_2, b_3, b_4)$, $C = (c_1, c_2, c_3, c_4)$ be three points not on a line in the 3-space. The line coordinates of AB and AC are

$$(a_1b_2 + a_2b_1, \cdots, a_3b_4 + a_4b_3) \text{ and } (a_1c_2 + a_2c_1, \cdots, a_3c_4 + a_4c_3).$$

These line coordinates considered as coordinates of two points in the 5-space determine a collinear point,

$$(a[b_2 + c_2] + a_2[b_1 + c_1], \cdots, a_3[b_4 + c_4] + a_4[b_3 + c_3])$$

whose coordinates are the line coordinates of the line joining A to the third point of BC, $(b_1 + c_1, b_2 + c_2, b_3 + c_3, b_4 + c_4)$. Hence the three points of a line in the 5-space which is entirely on the surface S represent a flat pencil, three lines in the 3-space which lie in a plane and pass through a point.

The seven lines in a plane of the 3-space are represented by the 7 points of a plane wholly on the surface S in the 5-space, since in the plane in the 3-space every line meets every other line. In a similar manner the lines through a point in the 3-space are represented by the points of a plane which is wholly on the surface S in the 5-space.

The tangent 4-space at a point of the surface S has already been shown to have 19 points in common with the surface S, representing the 18 lines which meet a given line in the 3-space, and the line itself.

Since through a given line in the 3-space there are 3 planes and through each point in a plane 3 lines, it follows that the 19 points which the tangent 4-space has in common with the surface S are upon 9 lines through the point of tangency.

Any 4-space passing through the point of tangency has in common with the tangent 4-space a tangent 3-space. These tangent 3-spaces are of two kinds. The first kind has in common with the surface S 11 points, which are upon 5 lines through the point of tangency, representing a degenerate congruence determined by two degenerate complexes whose axes have a point in common. The second kind has in common with the surface S 7 points, which are upon 3 lines through the point of tangency; these represent the 7 lines of a degenerate congruence with an axis.

The tangent 2-spaces are determined by two 4-spaces through the point of tangency and by the tangent 4-space. They are of 5 kinds. One and two have 7 points in common with the surface S and are on 3 lines through the point of tangency. They represent in the 3-space the lines of a plane and the lines through a point respectively. Three has 5 points in common with the surface S, which are on 2 lines through the point of tangency, and represent in the 3-space a line and a flat pencil through each of two points of the line determining two planes through the line. Four has three points of a line

in common with the surface S and represents a flat pencil of three lines in the 3-space. Five is a tangent plane which has but a single point in common with the surface S and represents a single line in the 3-space.

In the configuration of the 28 points off the surface S there are 70 planes, which have 6 points off the surface S. Two of these planes pass through each of the 35 points of the surface S. We have seen that the notation for the points of one of these planes is $(ab)(ac)(cb)(ad)(bd)(cd)$ and the notation for the plane itself may be taken as $(abcd)$.

If the heptads are denoted by the 8 letters a,b,c,d,e,f,g,h, it will be found that the 2 planes $(abcd)(efgh)$ have but 1 point in common with the surface S and that these common points are the same. Hence the points of the surface S can be denoted by a double 4 letter notation, which will also be the notation for the lines of the 3-space. On page 63 the line coordinates of all the lines are obtained from those of the 3 lines $(1,2,5)$, $(1,3,9)$, $(1,6,11)$, by use of the transformation T and its powers. The 8 letter notation for all the lines of the 3-space can be obtained from the notations for the 3 lines above, which are respectively $(1278 + 3456)$, $(1357 + 2468)$, $(1467 + 2358)$, by use of the transformation on the heptads, which is isomorphic with T, i.e., $(17654)(283)$.

The Determination of an Eight Letter Notation for the Points and Planes of the 3-space. A point of the 3-space is determined by the 7 lines which pass through it, and these correspond to the 7 points of a plane in the 5-space.

A plane of the 3-space is determined also by the 7 lines which lie in it, and these correspond in the 5-space to the 7 points of a plane. To each of the points of the 5-space there belongs a double four letter notation. Hence a point of the 3-space will be denoted by 14 sets of 4 numbers. The eight letter notation for the point $1 = (1,1,1,1)$ of the 3-space is $1278 + 1458 + 1234 + 1357 + 1256 + 1368 + 1467$, $3456 + 2367 + 5678 + 2468 + 3478 + 2457 + 2358$ and the notation for the point i is obtained from this by the ith power of T, where $T = (17654)(283)$.

These 14 sets of 4 numbers constitute the planes of a finite Euclidian geometry, where the numbers are considered as points; there are 4 points to a plane and 2 points to a line. This same Euclidian geometry can be obtained from the $PG(3,2)$ by striking out all the points of a plane, which will cut out 3 points from each plane. Another statement of the above is that the 14

sets of 4 numbers constitute a quadruple system,* for when any 3 numbers are given there is always a fourth determined.

The proof is as follows. When 3 numbers as a, b, c, are given this determines a line in the 5-space; this line determines a congruence in the 3-space whose 5 lines have the notation

$$abcd + efgh,\ abce + dfgh,\ abcf + degh,\ abcg + defh, \text{ and } abch + defg.$$

But through any point of the 3-space there is but one line belonging to a congruence, hence the fourth letter belonging to abc in the notation for this line is determined when this line is given.

The group, which leaves a plane of the 3-space or a point of the 3-space fixed, is of order $14 \cdot 12 \cdot 8 = 1344$ and this is the order of the group of the Euclidian geometry and of the quadruple systems on 8 letters.

There are 15 quadruple systems related to the planes of the 3-space and 15 related to the points of the 3-space. The points of the 3-space are transformed into the planes and vice versa by the odd transformations of the symmetric group on 8 letters, hence the 2 sets of 15 quadruple systems are conjugate under the symmetric group, while the members of each set of 15 are conjugate under the alternating group, which changes points into points and planes into planes.

The 8 letter notation for the points of the 3-space enables one to calculate the corresponding transformations of the $L.\,H.\,G.\ \overset{15}{\dfrac{|8}{2}}$ and the $G\overset{8}{\dfrac{|8}{2}}$ immediately. The method of passing from the linear homogeneous group to the alternating group has already been given on page 11. To pass in the reverse direction it is only necessary to determine into what points the vertices of the fundamental tetrahedron are transformed. This gives a method, which does not require the direct use of a table of corresponding transformations.†

Applications to Kirkman's School Girl Problem.‡ The problem is "to arrange 15 school girls in parties of 3 for 7 consecutive day's walks, so that no 2 girls may walk together more than once during the 7 days."

* E. H. Moore, loc. cit.

† Such a table is given by L. E. Dickson. *Mathematische Annalen*, vol. 54 (1901), page 564.

‡ See Ball, *Mathematical Recreations and Problems*, page 89, for numerous references to the problem.

A congruence of the 3-space as

5, 10, 15			1467 + 2385
1, 6, 11			1567 + 2384
2, 7, 12		or	4567 + 2381
3, 8, 13			1456 + 2387
4, 9, 14			1457 + 2386

(the latter in the 8 letter notation) is obtained by polarizing for the surface S with respect to the 3 points of a line in the 5-space which is wholly off the surface S, and taking the points of the 5-space which are common to the surface S and the 3 polar 4-spaces; these common points represent the lines of a congruence in the 3-space. The above congruence is obtained by polarizing with respect to the 3 points

(1,0,1,0,0,1)		28
(1,1,0,0,1,0)	or	38
(0,1,1,0,1,1)		23

(the latter in the 8 letter notation). The group of a congruence is thus composed of all members of the alternating group which permute 3 letters among themselves, hence is of the order $5 \cdot 3/2 = 360$. A day of the school girl problem is evidently a congruence.

A week's solution consists of 7 congruences which do not have a line in common. The 5 lines of a congruence can be written in the 8 letter notation $abcd + efgh$, $abce + dfgh$, $abcf + degh$, $abcg + defh$, and $a\ ch + defg$, i. e., they are given by $abcx + defgh/x$, $x = d, e, f, g, h$, and the congruence is determined by polarizing with respect to the 3 points ab, bc, ca, of a line in the 5-space. There are 2 types of congruences having a line in common with the above congruence, I, $abdx + cefgh/x$, $x = c, e, f, g, h$, obtained by polarizing with respect to the points ab, bd, da, and II, $abcdh/x + efgx$, $x = a, b, c, d, h$, obtained by polarizing with respect to the points ef, fg, ge. Hence in order that 2 congruences shall have a line in common, the lines in the 5-space with respect to which we polarize must either have a point in common as ab or they must not belong to the same heptad. Hence the school girl problem consists of finding 7 lines in the 5-space which do not intersect and such that any 2 lines always have a heptad in common.

The 8 heptads give a complete solution of the problem. We take 7 of the heptads as (1234567) and form a $PG(2,2)$ with heptads as points, and the following triads as lines:

123, 145, 167, 347, 246, 257, 356.

Each of the sets of 3 numbers determines a line in the 5-space which is wholly off the surface S and no 2 of the 7 lines have a point in common and each two have a heptad in common. These 7 lines in the 5-space thus determine a solution of the problem.

There are 30 $PG(2,2)$ related to 7 letters, and since there are 8 heptads there are $30 \cdot 8 = 240$ solutions of the school girl problem belonging to this geometry. The 30 $PG(2,2)$ belonging to a set of 7 heptads are conjugate under the symmetric group and there are 2 sets of 15 such that the members of each set are conjugate among themselves under the alternating group. Hence there are 2 sets of 120 solutions each, the solutions of each set are permuted among themselves by the projective collineation group * and one set is transformed into the other by a polarity.

If we apply the cyclic transformation (12345678) to the above $PG(2,2)$ we obtain 8 solutions, which do not have a congruence or day in common.† These 8 solutions embrace all 56 congruences. By a transformation of period 15 we can obtain 120 solutions from these 8 and by means of a duality all the 240 are obtained.

Each solution is invariant under a group of order 168, since it is a $PG(2,2)$ on 7 heptads. From this $PG(3,2)$ we derive 240 solutions. There are $2\,\underline{|15}/8$ equivalent spaces which are conjugate with this space, hence there are $240 \cdot 2 \cdot \underline{|15}/8$ solutions of the school girl problem to be obtained from these spaces. J. Power has shown that this is the number of possible solutions, hence each school girl solution is related to a $PG(3,2)$.‡

Application to the Equations of the Eighth and Lower Degrees. A particular expression,§

$$v = x_1x_2x_7x_8 + x_1x_4x_5x_8 + x_1x_2x_3x_4 + x_1x_3x_5x_7 + x_1x_2x_5x_6 + x_1x_3x_6x_8 + x_1x_4x_6x_7$$
$$+ x_3x_4x_5x_6 + x_2x_3x_6x_7 + x_5x_6x_7x_8 + x_2x_4x_6x_8 + x_3x_4x_7x_8 + x_2x_4x_5x_7 + x_2x_3x_5x_8,$$

---

* E. H. Moore, loc. cit., page 441.

† E. H. Moore, loc. cit., page 443.

‡ J. Power, On the Problem of the Fifteen School Girls. *Quarterly Journal of Mathematics*, vol. 8 (1867), pp. 236-251.

§ Mathieu, *Journal de mathématiques pures et appliqués*, vol. 6 (1861), pp. 241—323. See page 291.

which is the notation for a point or a plane in the 3-space and is invariant under a group of substitutions of order $14 \cdot 12 \cdot 8 = 1344$, has been used to reduce the general equation of the eighth degree to a special one whose Galois group is of order 1344.* The expression v has 15 conjugates under the alternating group and hence is the root of an equation of the fifteenth degree whose coefficients can be rationally expressed in terms of the coefficients of the original equation and of the square root of the discriminant. On adjoining a root v of this equation of the fifteenth degree and adjoining the square root of the discriminant, the general equation of eight degree reduces to a particular one with the Galois group of order 1344.

The equation of the seventh degree may be considered by allowing one of the heptads to be fixed. A function which plays a similar role to the one above for the equation of the seventh degree is

$$v = x_1x_5x_7 + x_2x_6x_1 + x_3x_7x_2 + x_4x_1x_3 + x_5x_2x_4 + x_6x_3x_5 + x_7x_4x_6.$$

This is the notation for a school girl solution and as has been shown is invariant under a group of order 168. Thus v has 15 conjugates under the alternating group. Hence on adjoining a root of an equation of the fifteenth degree and adjoining also the square root of the discriminant, the general equation of the seventh degree reduces to a special equation whose Galois group is of order 168.†

New Haven, Conn.,

April, 1909.

* H. Weber, *Lehrbuch der Algebra*, vol. 2, page 377.

† H. Weber, loc. cit, page 540.

THE GEODESIC LINES ON THE HELICOID

By S. E. Rasor

Introduction. To find the geodesics on the helicoid we will first obtain them for a surface to which the helicoid is applicable and on which they are more easily obtained. Since lengths are preserved in the application of surfaces and since geodesics are shortest lines they must correspond on two applicable surfaces. Or, analytically, the geodesics will correspond on two applicable surfaces since only E, F, and G and their derivatives enter the differential equation of the geodesics.

It has been shown* that the helicoid whose equations are

$$x = u' \cos v', \quad y = u' \sin v', \quad z = av',$$

u' being the radius vector and v' the angle made by the x-axis and the projection of the radius vector on the xy-plane, is applicable to the catenoid whose equations are

$$x = u \cos v, \quad y = u \sin v, \quad z = a \text{ arc cosh } (u/a),$$

$$u = \frac{a}{2}\left(e^{z/a} + e^{-z/a}\right),$$

where u and v are defined as for the helicoid and where $u \geqq a$.† The two surfaces are applied when we choose as corresponding points those for which

$$u^2 = a^2 + u'^2, \quad v = v'.$$

The correspondence of the two surfaces when one is thus applied to the other has also been exhibited,‡ viz., the smallest circle of the catenoid corresponds to the axis of the helicoid, other circles corresponding to helices; the meridians of the catenoid correspond to the straight line generators of the helicoid.

---

* Darboux, *Théorie générale des surfaces*, vol. 1, pp. 77, 82.

† The discussion is limited throughout to real points and real geodesics.

‡ Darboux, l. c., vol. 1, p. 83.

The Geodesics on the Catenoid.* The differential equation of the geodesics on a surface of revolution† becomes

$$(1) \qquad u^2 \frac{dv}{ds} = c, \quad ds^2 = dx^2 + dy^2 + dz^2.$$

If $c = 0$ in this equation we obtain, since $u = 0$ is not possible, $v =$ constant, so that the meridians on the catenoid are geodesics. Of the parallel circles only $u = a$ is a geodesic since it is the only one for which the principal normal is normal to the surface, this being a necessary and sufficient condition.

By substituting in ds in equation (1) the values of dx, dy, dz from the equations of the catenoid, we obtain

$$(2) \qquad v = c \int \frac{du}{\sqrt{(u^2 - a^2)(u^2 - c^2)}}.$$

There are three cases of this equation to discuss, viz.,

$$c > a, \quad c < a, \quad c = a,$$

in which only positive values of c will be considered since changing the sign of c only changes the sign of v in (1).

First case: $c > a$. Since $u^2 \geqq a^2$, it follows that $u^2 > c^2$ if the integral be real. Substituting $u = c/\sin\phi$, equation (2) reduces to

$$(3) \qquad v = \int_0^{\phi} \frac{d\phi}{\sqrt{1 - k^2 . \sin^2\phi}} + m,$$

in which $k = a/c$ and m is any constant; m can be chosen equal to zero without loss of generality since the geodesics for the different values of m can be made to coincide by a rotation about the z-axis. Thus

$$u = \frac{c}{\sin\phi}, \quad v = F(k, \phi), \qquad 0 \leqq \phi \leqq \pi$$

are the equations of one branch of any geodesic of the first kind. But $\sin\phi = c/u = \cos\alpha$ from Clairaut's equation,‡ proved by writing (1) in the

* Cf. Darboux, l. c., vol. 3, p. 4, for a brief discussion of the geodesics on a surface of revolution having meridians extending to infinity and a parallel of minimum radius.

† Knoblauch, *Theorie der krummen Flächen*, pp. 147, 148.

‡ Knoblauch, *l. c.*, p. 148.

form $u \cdot \dfrac{udv}{ds} = c$, in which a is the angle between an arc of the geodesic and the arc of the circle. It follows that

$$\phi = 2n\pi + \tfrac{1}{2}\pi \pm a,$$

where n is any integer.

But on the circle $u = c$, supposed above the xy-plane, $\sin \phi = \pm 1$ and therefore $a = n\pi$. The geodesic is therefore perpendicular to the meridian at the point where it is crossed by the circle $u = c$. Let u be positive and increase from c. Suppose ϕ to start from $\pi/2$ and to approach zero; then u increases indefinitely, and the geodesic approaches asymptotically the meridian $v = 0$. Considerations of symmetry at once show that the geodesic is symmetrical with respect to the meridian plane, $v = F(k, \pi/2)$, since u remains the same for both $\pi/2 + \phi$ and $\pi/2 - \phi$. A branch of the geodesic for a given c is thus enclosed between the planes $v = F(k, 0) = 0$ and $v = 2\,F(k, \pi/2) = F(k, \pi)$ and lies above the circle $u = c$ since $u > c$ along the geodesic. As ϕ increases from π to 2π another branch, equal to the one just described, is obtained which is tangent to the circle $u = c$ at $v = F(k, 3\pi/2)$. The equations of this branch may be written, since u and c are considered positive,

$$-u = c \sin \phi, \quad v = F(k, \phi), \qquad \pi \leqq \phi \leqq 2\pi.$$

This process may be continued indefinitely.

If $2F(k, \pi/2)$ is not commensurable with 2π the branches of the geodesic for a given c are repeated in endless procession without returning to the starting point. If the ratio of $2F(k, \pi/2)$ to 2π is equal to p/q where p and q are integral, then while v describes p complete revolutions the geodesic consisting of q branches returns to the starting point and may be said to be "closed" although each of its branches runs to infinity. But $2F(k, \pi/2)$ is a continuous function of k, varying monotonically from π to ∞ as k varies from 0 to 1. There must therefore be infinitely many values of k for which $2F(k, \pi/2)$ is commensurable with 2π. Consequently "closed" geodesics exist on every catenoid.

Moreover, the angle $2F(k, \pi/2)$ becomes infinite as c approaches a and k approaches unity, so that as the circle, $u = c$, approaches the smallest one, $u = a$, the geodesic winds about the catenoid an increasing number of times before it becomes tangent to $u = c$. But as c increases, $2F(k, \pi/2)$ decreases

and approaches its minimum π as c becomes infinite and k approaches zero. The limiting "closed" geodesic (not a circle) therefore winds about the catenoid just once and goes to infinity once.

SECOND CASE: $c < a$. By the substitution $u = a/\sin\phi$, equation (2) reduces to

$$v = k\int_0^{\phi}\frac{d\phi}{\sqrt{1-k^2\sin^2\phi}} + m, \tag{4}$$

where $k = c/a$ and, as before, m may be taken equal to zero, and where we study the behaviour of u and v for values of ϕ in the interval $0 \leqq \phi \leqq \pi$. It is to be noticed that ϕ no longer has the geometrical interpretation of case one. Thus, for this branch,

$$u = \frac{a}{\sin\phi}, \quad v = k\ F(k,\ \phi), \qquad 0 \leqq \phi \leqq \pi$$

are the equations of the geodesic.

If $\phi = \pi/2$, then $u = a$ and $\alpha = \text{arc}\cos\ (c/a)$ since $\cos\ \alpha = c/u$. As ϕ increases to π, u increases without limit, and α increases to $\pi/2$, so that the geodesic approaches a meridian asymptotically. Meanwhile v is increased by $kF(k,\ \pi/2)$.

But the surface is symmetrical with respect to each meridian plane, and also with respect to the plane of the circle, $u = a$. As v, therefore, varies from 0 to $2\ kF(k,\ \pi/2) = kF(k,\ \pi)$, a branch of the curve starts at infinity on the upper or lower part of the surface at the plane $v = 0$, crosses the circle $u = a$ at an angle $\alpha = \text{arc}\cos\ (c/a)$ at the point P in the plane $v = kF(k,\ \pi/2)$, and passes to infinity on the lower or upper part of the surface where it approaches asymptotically the meridian $v = kF(k,\ \pi)$. There is a pencil of geodesics through this point P, each one intersecting $u = a$ in its own angle corresponding to the values of c. As v increases again by $kF(k,\ \pi)$, another branch of the geodesic is obtained crossing the circle $u = a$ at the plane $v = k\ F(k,\ 3\pi/2)$; so we may proceed indefinitely.

But as c approaches a and k approaches unity, $kF(k,\ \pi)$ becomes infinite and the geodesic winds about the catenoid increasingly often before crossing $u = a$, and the angle of intersection with $u = a$ approaches zero. As c and

k approach zero, $kF(k, \pi)$ approaches zero, and thus the limiting case of this class of geodesics is a meridian. If $2kF(k, \pi/2)$ is commensurable with 2π, the geodesic returns to its starting point and may be called "closed", by which is meant however only that for a given c the geodesic consists of a finite number of branches. There are an infinite number of values of k for which this is true. If it is not commensurable with 2π, the branches are repeated in endless procession and never return to the starting point.

THIRD CASE: $c = a$. If in equation (4) we put $k = c/a = 1$, the equation of the geodesics reduces to

$$(5) \qquad u = \frac{a}{\sin\phi}, \quad v = \int_0^\phi \frac{d\phi}{\cos\phi} = \log\ \tan\left(\frac{\pi}{4} + \frac{\phi}{2}\right).$$

But
$$\tanh v = \frac{e^v - e^{-v}}{e^v + e^{-v}}$$

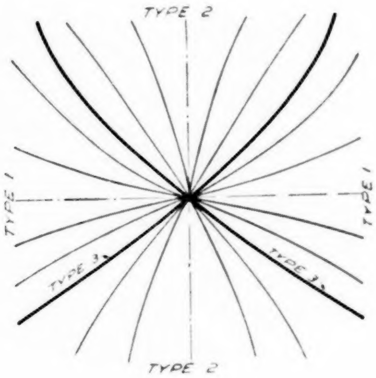
and, by substituting for v from (5),

$$\tanh v = \frac{2\tan\dfrac{\phi}{2}}{\sec^2\dfrac{\phi}{2}} = \text{in}\ \phi = \frac{a}{u}$$

Therefore along the geodesic, $u = a \coth v = a\dfrac{e^v + e^{-v}}{e^v - e^{-v}}$. This equation in the polar coordinates u, v in the xy-plane represents a cylinder whose intersection with the catenoid will give the geodesics for $c = a$. As v approaches zero, u becomes infinite, and as v increases indefinitely u decreases to a. The geodesic thus starts tangent asymptotically to the meridian, $v = 0$, comes into the finite part of the surface as v increases, and with constantly decreasing radius vector winds about the circle $u = a$ as v increases indefinitely. It is the limiting case of the geodesics of classes one and two.

Through a given point on the surface there will then pass a pencil of geodesics of type one ($c > a$), *a pencil of geodesics of type two* ($c < a$), *and two geodesics of type three* ($c = a$), *so situated that the two geodesics of type three separate those of type one from those of type two.* This follows immediately from Clairaut's equation, $u \cos\alpha = c$. For a given point, u is

fixed and as c decreases continuously from u to zero through a, the angle α increases continuously from zero to $\pi/2$.



The Geodesics on the Helicoid. As has been said, the correspondence between the helicoid

$$x = u' \cos v', \qquad y = u' \sin v', \qquad z = av'$$

and the catenoid

$$x = u \cos v, \qquad y = u \sin v, \qquad z = a \text{ arc cosh } (u/a)$$

in the application of these surfaces is given by

$$u^2 = a^2 + u'^2, \qquad v = v'.$$

To a meridian $v = v_0 + 2n\pi$ of the catenoid correspond parallel straight line generators

$$x = u' \cos v_0, \qquad y = u' \sin v_0, \qquad z = a(v_0 + 2n\pi)$$

of the helicoid.

To the minimum parallel circle, $u = a$, of the catenoid, corresponds $u' = 0$, the axis of the helicoid, but to each point of the circle correspond on this axis an infinite number of points, $2\pi a$ units apart.

To any other parallel of the catenoid $u = b > a$ corresponds a helix $u' = \sqrt{b^2 - a^2}$ on the catenoid, but to each point of the circle correspond an infinite number of points on the helix lying in a line parallel to the z-axis (vertical) at a distance $2\pi a$ units apart. The correspondence of points in the

two surfaces may be made more precise by considering only those points of the catenoid on one side of the minimum parallel, for example only points for which $z \geqq 0$, and by supposing $u > 0$; then to each pair of values of u and v corresponds a single point of the surface, and to each point of the surface, only one value of u, and values of v differing by multiples of 2π.

If in the equations of the helicoid we suppose $u' > 0$, we shall have a single point corresponding to each pair of values of u' and v', and conversely, to each point of the surface one value of u', and one value of v'. So that both u and u' being positive we have to one point of the helicoid a single corresponding point of the catenoid, but to each point of the latter correspond points of the helicoid in the same vertical line, $2\pi a$ units apart. It is on the basis of such a correspondence that the equations of the geodesics of the helicoid are given in terms of elliptic functions. But in the discussion of the geometrical properties of the geodesics of the second class—those crossing the minimum parallel of the catenoid and the axis of the helicoid—the continuations of these lines have been included, namely points of the catenoid for which z is negative, and points of the helicoid for which u' is negative.

If the double signs are omitted from those equations they will indeed be still the equations of geodesic lines, but the correspondence would perhaps be not so clearly exhibited.

Each branch of a geodesic on the catenoid for the first case, $c > a$, was found between meridians making an angle of $F(k, \pi)$ with each other. They are tangent to a certain circle, $u = c$, at a point half way between two such meridians. A corresponding helicoid geodesic lies between planes which pass through the axis and differ in angle by $F(k, \pi)$, as appears from the geometrical meanings of v and v'. The geodesic approaches asymptotically the generators of the helicoid which lie in these planes and touches a certain helix corresponding to $u = c$. There is a pencil of geodesics of this type through each point of the surface.

If $F(k, \pi)$ is commensurable with 2π, other branches of the same geodesic for the same c are tangent to the same helix vertically over the first point of tangency; in other words there are equal geodesics, tangent to the same helix at points in a vertical line, which approach asymptotically the same generator. If not commensurable this is not the case. But at the limit, when $c = a$ and $F(k, \pi)$ is infinite, the geodesic of the third class winds about the helicoid indefinitely and only approaches the axis $u' = 0$. Moreover as c is indefinitely

increased, $F(k, \pi)$, the angle between the asymptotic generators, approaches its minimum value π.

The equations of these geodesics may be written in terms of elliptic functions. The equations of the corresponding catenoid geodesics are

$$\pm u \sin \phi = c, \qquad v = F(k, \phi), \qquad k = \frac{a}{c}, \qquad \phi = amv.$$

Therefore

$$\pm u \operatorname{sn} v = c, \quad \text{or} \quad \sqrt{u'^2 + a^2} \cdot \operatorname{sn} v = c = a/k.$$

Hence

$$u'^2 \cdot \operatorname{sn}^2 v = \frac{a^2}{k^2}(1 - k^2 \operatorname{sn}^2 v), \quad \text{or} \quad u' = \pm \frac{a \operatorname{dn} v'}{k \operatorname{sn} v'}$$

since $v = v'$. Thus the required equations are

$$x = u' \cos v' = \pm \frac{a \operatorname{dn} v' \cdot \cos v'}{k \operatorname{sn} v'},$$

$$y = u' \sin v' = \pm \frac{a \operatorname{dn} v' \cdot \sin v'}{k \operatorname{sn} v'},$$

$$z = c v'.$$

We suppose u' positive throughout, so that the signs of x and y, in the equations above, are to be respectively those of $\cos v'$ and $\sin v'$.

For the second case, $c < a$, a branch of the catenoid geodesic crosses the circle $u = a$ midway between two meridians making an angle of $kF(k, \pi)$ with each other and is asymptotic to these meridians. Hence the branch of the corresponding helicoid geodesic crosses the axis at an angle $\alpha = \operatorname{arc\,cos} (c/a)$ midway between two generators asymptotic to the geodesic and making an angle of $v' = kF(k, \pi)$ with each other. If $kF(k, \pi)$ is commensurable with 2π there are equal branches of geodesics issuing from points on the axis distant $2n\pi a$ with parallel tangents at those points with a common asymptotic generator. If $kF(k, \pi)$ is not commensurable with 2π, this is not the case. There are an infinite number of values of k for which $kF(k, \pi)$ is commensurable with 2π.

The equation of the geodesics of this class may be written as follows:

$$x = u' \cos v' = \pm \frac{a \cos v'}{\operatorname{tn}(v'/k)},$$

$$y = u' \sin v' = \pm \frac{a \sin v'}{\operatorname{tn}(v'/k)},$$

$$z = av',$$

where $u' = \dfrac{a}{\operatorname{tn}(v'/k)}$, and $k = a/c$ from the equations of the corresponding catenoid geodesics. As in the first case, the sign of u' being throughout taken as positive, the signs of x and y are respectively those of $\cos v'$ and $\sin v'$.

The catenoid geodesics for the third case $c = a$ were found when projected on the xy-plane to be

$$u \tanh v = a,$$

which for the helicoid reduces to

$$u'(e^{v'} - e^{-v'}) = 2a,$$

since $u^2 = u'^2 + a^2$. If v' approaches zero, u' increases indefinitely; and as v' increases u' decreases and approaches zero. Therefore, as v' increases from 0 to 2π, the corresponding geodesic starting at an infinite distance traverses the entire helicoid and at each period of 2π lies closer to the axis.

The equations of the geodesics of this class may be written in the form

$$x = u' \cos v' = a \operatorname{cosech} v' \cos v',$$

$$y = u' \sin v' = a \operatorname{cosech} v' \sin v',$$

$$z = av',$$

where $u' = a \operatorname{cosech} v'$. Through each point of the helicoid, not on the axis, pass two geodesics of this class, separating, as in the catenoid, those of the first and second classes which pass through that point.

Ohio State University,
October, 1909.

CUBIC CONGRUENCES WITH THREE REAL ROOTS

By Edward B. Escott

Gauss has shown that the complete solution of the equation

$$x^n - 1 = 0$$

where n is prime, is found by solving some auxiliary equations whose degrees are the factors of $n - 1$. These equations are called cyclotomic equations. Gauss showed that these equations are irreducible.

Consider, for example, the equation

$$x^7 - 1 = 0. \tag{1}$$

Calling one of its complex roots ω, the remaining roots are ω^2, ω^3, ω^4, ω^5, ω^6, $\omega^7 = 1$.

Since the sum of the roots of (1) is zero, we have

$$1 + \omega + \omega^2 + \omega^3 + \omega^4 + \omega^5 + \omega^6 = 0. \tag{2}$$

Arrange the roots so that the exponents are in geometrical progression. Several groupings are possible :

$$\begin{aligned} &\omega,\ \omega^2,\ \omega^4,\ \omega^8\ (=\omega),\ \cdots \\ &\omega^3,\ \omega^6,\ \omega^{12}\ \ (=\omega^5),\ \omega^{24}\ (=\omega^3),\ \cdots \end{aligned} \tag{3}$$

i. e., two groups of three each : or

$$\begin{aligned} &\omega,\ \omega^6,\ \omega^{36}\ (=\omega),\ \cdots \\ &\omega^2,\ \omega^{12}\ (=\omega^5),\ \omega^{72}\ (=\omega^2),\ \cdots \\ &\omega^3,\ \omega^{18}\ (=\omega^4),\ \omega^{108}\ (=\omega^3),\ \cdots \end{aligned} \tag{4}$$

i. e., three groups of two each.

If we take the sum of the roots in the same row in (3) for roots of a new equation, i. e.

$$a = \omega + \omega^2 + \omega^4, \qquad \beta = \omega^3 + \omega^6 + \omega^5, \tag{5}$$

we shall have

$$x^2 + x + 2 = 0 \tag{6}$$

with the roots a and β.

Similarly, in (4) if we put

$$a = \omega + \omega^6, \qquad \beta = \omega^2 + \omega^5, \qquad \gamma = \omega^3 + \omega^4, \tag{7}$$

the equation whose roots are a, β, γ is

$$x^3 + x^2 - 2x - 1 = 0. \tag{8}$$

Equations (6) and (8) are the cyclotomic equations for the division of the circle into seven parts.*

From (7), $a^2 = \omega^2 + 2\omega^7 + \omega^{12} = \omega^2 + 2 + \omega^5 = \beta + 2$; also $\beta^2 = \gamma + 2$, and $\gamma^2 = a + 2$.

From these relations, we see that all the roots of (8) can be obtained from any one root, and since (8) must have one real root, all of its roots must be real.

In order to apply these results to cubic congruences with modulus an odd prime, let us summarize briefly a few of the properties of these congruences.

A congruence cannot have more roots than its degree. The degree of a congruence (mod p) can always be reduced to $p - 2$ by Fermat's Theorem.

$$x^{p-1} \equiv 1 \pmod{p}, \qquad x \not\equiv 0 \pmod{p}.$$

A congruence of the first degree always has a root.

A congruence of the second degree has two roots or none.

A congruence of the third degree has (a) three roots, (b) one root, or (c) no roots.

The condition that a cubic congruence shall have three roots is rather involved, so it seems of interest to consider a large class of congruences which have three roots (when they have one).

An example is obtained at once from (8).

If the congruence

$$x^3 + x^2 - 2x - 1 \equiv 0 \qquad (\text{mod } p) \tag{9}$$

has a root a, it is evident that it has also the roots $\beta = a^2 - 2$ and $\gamma = \beta^2 - 2$.

* For details see Mathews' *Theory of Numbers*.

Let us consider the problem, to find the most general irreducible cubic equation in which each root is a rational integral function of another root, i. e.,

$$\beta = f(\alpha), \qquad \gamma = f(\beta), \qquad \alpha = f(\gamma),$$

and let the equation whose roots are α, β, γ be

$$x^3 + ax^2 + bx + c = 0\,; \tag{10}$$

then by using the relation

$$\alpha^3 + a\alpha^2 + b\alpha + c = 0$$

and similar ones for β and γ, we can replace $f(\alpha)$ by a function of the second degree.

Consider first the case where the roots have the relations

$$\beta = \alpha^2 - n, \qquad \gamma = \beta^2 - n, \qquad \alpha = \gamma^2 - n, \tag{11}$$

and let α, β, γ be roots of (10).

Form the equation whose roots are the squares of the roots of the given equation, by transposing the terms of even degree to the second member, squaring both members, and replacing x^2 by x. We have

$$x^3 + (2b - a_2)\, x^2 + (b^2 - 2ac)\, x - c^2 = 0. \tag{12}$$

If we increase the roots of the given equation (10) by n, we shall have

$$x^3 + (-3n + a)\, x^2 + (3n^2 - 2an + b)\, x + (-n^3 + an^2 - bn + c) = 0. \tag{13}$$

These two equations must be identical. Equating coefficients,

$$2b - a^2 = -3n + a, \tag{14}$$

$$b^2 - 2ac = 3n^2 - 2an + b, \tag{15}$$

$$c^2 = n^3 - an^2 + bn - c. \tag{16}$$

From (14)

$$n = \tfrac{1}{3}\,(a^2 + a - 2b). \tag{17}$$

Substituting in (15) and solving for c,

$$c = -\frac{1}{6a}\,(a^4 - 4a^2b - a^2 + b^2 + 3b). \tag{18}$$

Substituting these values of n and c in (16) and arranging with reference to b, we have

$$\begin{aligned}&3b^4 + (8a^2 + 18)\, b^3 + (6a^4 - 6a^2 - 18a + 27)\, b^2 \\ &- (18a^4 - 36a^3 + 18a^2 + 54a)\, b - (a^8 - 6a^6 + 10a^5 - 3a - 18a^3) = 0.\end{aligned} \tag{19}$$

We can tell some of the roots of (19) at once. For example, in (11) if $\gamma = \beta = \alpha$, we have $n = \alpha^2 - \alpha$; and since (10) becomes $(x - \alpha)^3 = 0$, we have $\alpha = -a/3$; whence $n = (a^2 + 3a)/9$ and from (14)

$$b = \frac{a^2}{3}. \tag{20}$$

This is one root of (19).

In (11) instead of α, β, and γ being different, one root might be repeated. Instead of (11) we would have

$$\alpha = \alpha^2 - n, \qquad \beta = \beta^2 - n, \qquad (\beta \neq \alpha). \tag{21}$$

Subtracting, $\alpha - \beta = \alpha^2 - \beta^2$, i. e., $1 = \alpha + \beta$, whence $\beta = -\alpha + 1$.

Equation (10) becomes

$$(x + \alpha)^2\, (x + \alpha - 1) = 0, \tag{22}$$

and comparing coefficients with (10) we find $a = -\alpha - 1$, $b = -\alpha^2 + 2\alpha$, whence $\alpha = -a - 1$, and

$$b = -(a^2 + 4a + 3). \tag{23}$$

This gives another root of (19).

In place of (11) if we had taken the relations

$$\alpha = \alpha^2 - n, \qquad \beta = \gamma^2 - n, \qquad \gamma = \beta^2 - n \tag{24}$$

we would have the same equations (14), (15), and (16). From the last two equations $\beta - \gamma = \gamma^2 - \beta^2$, and since $\beta \neq \gamma$, $1 = -\gamma - \beta$, i. e., $\gamma = -\beta - 1$. Substituting in the second or third relation of (24), we have

$$\beta^2 + \beta - n + 1 = 0. \tag{25}$$

Therefore, α, β, γ are roots of the equation

$$(x - \alpha)(x^2 + x - n + 1) = 0. \tag{26}$$

Equating coefficients of (26) and (10), we find $-\alpha + 1 = a$, and $-\alpha - n + 1 = b$, whence $\alpha = -a + 1$; and since $n = \alpha^2 - \alpha = a^2 - a$, we have

$$b = -(a^2 - 2a). \tag{27}$$

This gives a third root of (19).

Therefore, the remaining root of (19), and the only one which gives for (10) an irreducible equation, is

$$b = -(a^2 - 2a + 3). \tag{28}$$

From (17) and (18) we have, with this value of b,

$$c = -(a^3 - 2a^2 + 3a - 1),$$

$$n = a^2 - a + 2.$$

Then the equation

$$x^3 + ax^2 - (a \quad - 2a + 3)\, x - (a \quad - 2a^2 + 3a - 1) = 0 \tag{29}$$

has its roots α, β, γ connected by the relations

$$\begin{aligned} \beta &= \alpha^2 - (a^2 - a + 2), \\ \gamma &= \beta^2 - (a^2 - a + 2), \\ \alpha &= \gamma^2 - (a^2 - a + 2), \end{aligned} \tag{30}$$

and since it has one real root, all its roots are real.

The application to cubic congruences is immediate. We have the theorem:

The congruence

$$x^3 + ax^2 - (a^2 - 2a + 3)\, x - (a^3 - 2a^2 + 3a - 1) \equiv 0 \pmod{p}$$

has three roots (when it has any), the relations between the roots being given in equations (30).*

* We will have the same relations (30) between the roots if we replace a by $-(a - 1)$ This gives two irreducible cubic congruences having the same relations between their roots.

To find irreducible cubic equations or congruences, having between their roots a more general relation than (11), namely

$$\begin{aligned} \beta &= \alpha^2 + k\alpha + l, \\ \gamma &= \beta^2 + k\beta + l, \\ \alpha &= \gamma^2 + k\gamma + l, \end{aligned} \tag{31}$$

we could use the preceding method, or we can obtain the results from those already obtained, as follows:

Equations (31) may be written,

$$\begin{aligned} \beta + \frac{k}{2} &= \left(\alpha + \frac{k}{2}\right)^2 - \frac{k^2 - 2k - 4l}{4}, \\ \gamma + \frac{k}{2} &= \left(\beta + \frac{k}{2}\right) - \frac{k^2 - 2k - 4l}{4}, \\ \alpha + \frac{k}{2} &= \left(\gamma + \frac{k}{2}\right) - \frac{k^2 - 2k - 4l}{4}. \end{aligned} \tag{32}$$

Increase the roots of (10) by $k/2$, and we have

$$x^3 + \left(a - \frac{3k}{2}\right)x^2 + \left(b - ak + \frac{3k^2}{4}\right)x + \left(c - \frac{bk}{2} + \frac{ak^2}{4} - \frac{k^3}{8}\right) = 0. \tag{33}$$

Substitute in the results previously found the coefficients of (33) in place of a, b, and c, and $(k^2 - 2k - 4l)/4$ in place of n; we have, then :—

The congruence

$$x^3 + ax^2 + bx + c \equiv 0 \pmod{p}$$

has three roots (when it has any), the relations between the roots being

$$\begin{aligned} \beta &= \alpha^2 + k\alpha + l, \\ \gamma &= \beta^2 + k\beta + l, \\ \alpha &= \gamma^2 + k\gamma + l, \end{aligned}$$

where

$$\begin{aligned} b &= -\,(a^2 - 4ak - 2a + 3k^2 + 3k + 3), \\ c &= -\,(a^3 - 4a^2k - 2a^2 + 5ak^2 + 5ak + 3a - 2k^3 - 3k^2 - 3k - 1), \\ l &= -\,(a^2 - 3ak - a + 2k^2 + 2k + 2).^* \end{aligned}$$

* We will have the same relations between the roots if we replace a by $-a + 3k + 1$. This gives two irreducible cubic congruences having the same relations between the roots.

Finally, the equation whose roots have the most general relation of the second degree,

$$\begin{aligned} \beta &= f\alpha^2 + g\alpha + h, \\ \gamma &= f\beta^2 + g\beta + h, \\ \alpha &= f\gamma^2 + g\gamma + h, \end{aligned} \tag{34}$$

can easily be found from the preceding. Equation (34) may be written in the form

$$\begin{aligned} f\beta &= (f\alpha)^2 + g(f\alpha) + fh, \\ f\gamma &= (f\beta)^2 + g(f\beta) + fh, \\ f\alpha &= (f\gamma)^2 + g(f\gamma) + fh, \end{aligned} \tag{35}$$

which is like (31) with $f\alpha$, $f\beta$, $f\gamma$ in place of α, β, γ; g in place of k; and fh in place of l.

If α, β, γ are roots of

$$x^3 + ax^2 + bx + c = 0,$$

$f\alpha$, $f\beta$, $f\gamma$ are roots of

$$x^3 + afx^2 + bf^2x + cf^3 = 0.$$

The corresponding theorem is:

The congruence

$$x^3 + ax^2 + bx + c \equiv 0 \pmod{p}$$

has three roots (when it has any), the relation between the roots being

$$\begin{aligned} \beta &= f\alpha^2 + g\alpha + h, \\ \gamma &= f\beta^2 + g\beta + h, \\ \alpha &= f\gamma^2 + g\gamma + h, \end{aligned}$$

where

$$\begin{aligned} bf^2 &= -(a^2f^2 - 4afg - 2af + 3g^2 + 3g + 3), \\ cf^3 &= -(a^3f^3 - 4a^2f^2g - 2a^2f^2 + 5afg^2 + 5afg + 3af - 2g^3 - 3g^2 - 3g - 1), \\ hf &= -(a^2f^2 - 3afg - af + 2g^2 + 2g + 2). \end{aligned}$$

UNIVERSITY OF MICHIGAN
ANN ARBOR.

A GENERALIZATION OF THE GAME CALLED NIM

By E. H. Moore

In the third volume of the second series of the Annals of Mathematics Professor Bouton described and gave the complete mathematical theory of a known game for which he proposed the name *Nim*.

I propose to describe a generalization of Nim, which may be called Nim_k, read *Nim index k*. Here k is any positive integer, and the game Nim_1 is the original game Nim. Nim_k has likewise a complete mathematical theory which I shall content myself with formulating.

Description of the Game Nim_k. There are two players A and B and an assortment of objects of any kind, say counters. The dealer A takes as many counters as he wishes and separates them at will into any number ($\geqq 1$) of piles. The players draw alternately from this deal of say n piles, B drawing first; the player drawing the last counter (or counters) wins. In each draw the player must draw one or more counters from some one pile and he may draw at will from any number of piles not to exceed k. (Thus, in Nim_1 each draw is from one pile.)

Mathematical Theory of the Game Nim_k. It is clear that, if A deals to B fewer than $k+1$ piles, B may win on the first draw by drawing all the counters. Such a deal is an *unsafe combination* (to adopt a term used by Bouton) for A to deal to B. There are in fact two kinds of combinations: *safe* and *unsafe combinations*, the fundamental properties being that *every unsafe combination by a suitable draw may be made safe*, while *every safe combination by every draw is made unsafe*. Thus, if A deals a safe combination to B, B by drawing cannot avoid making it unsafe, A by drawing suitably makes it again safe, and so on until finally B is obliged to reduce the number of piles below $k+1$, when A wins. On the other hand, if A deals an unsafe combination to B, B by drawing suitably makes it safe, and then the game proceeds as before, until B finally wins.

Formula for safe combinations. Let the combination be of n piles containing respectively $c_1, c_2, \cdots, c_n$ counters.

Represent these n numbers

$$c_i \qquad (i = 1, \cdots, n)$$

in the binary scale of notation, i. e., determine integers

$$c_{ij} \qquad \begin{pmatrix} i = 1, \cdots, n \\ j = 0, 1, \cdots \end{pmatrix}$$

each 0 or 1, in such a way that

$$c_i = c_{i0} + c_{i1}\, 2^1 + c_{i2}\, 2^2 + \cdots + c_{ij}\, 2^j + \cdots \qquad (i = 1, 2, \cdots, n).$$

These integers c_{ij} are uniquely determinable. Then the combination is safe if and only if

$$\sum_{i=1}^{i=n} c_{ij} \equiv 0 \pmod{k+1} \qquad (j = 0, 1, 2, \cdots),$$

i. e., if and only if for every place j the sum of the n digits $c_{ij} (i = 1, \cdots, n)$ is exactly divisible by $k + 1$.

This definition and the theory as well as the game Nim_k are generalizations to $k = k$ from the case $k = 1$ of Nim.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

A SIMPLE METHOD FOR GRAPHICALLY OBTAINING THE COMPLEX ROOTS OF A CUBIC EQUATION

By Rutherford E. Gleason

Let $a \pm b\sqrt{-1}$ and c denote the roots of the equation

$$x^3 + px^2 + qx + r = 0.$$

Plot the curve $x^3 + px^2 + qx + r = y$. The curve cuts the x-axis at $P \equiv (c, 0)$. If a straight edge be revolved about P as a pivot until it touches the curve as a tangent at T, the abscissa of T will be a, and the slope of the tangent will be b^2. Thus the real part, a, of the complex roots may be found immediately, and the coefficient of the imaginary unit from the fact that $b^2(a - c)$ and the ordinate of T are equal.

This is easily shown as follows:

$$x^3 + px^2 + qx + r = y. \tag{1}$$

$$\frac{dy}{dx} = 3x^2 + 2px + q.$$

The tangent at the point (x_1, y_1) is therefore

$$y - y_1 = (3x_1^2 + 2px_1 + q)(x - x_1). \tag{2}$$

Since $a \pm b\sqrt{-1}$ and c are the roots of $x^3 + px^2 + qx + r = 0$,

$$p = -2a - c, \qquad q = a^2 + 2ac + b^2, \qquad r = -a^2c - b^2c.$$

Putting $x = a$ and substituting the values of p, q, r in (1),

$$y = b^2(a - c).$$

Putting $x_1 = a$, $y_1 = b^2(a - c)$, and substituting the values of p and q in (2), we have

$$y = b^2(x - c)$$

as the tangent at $T \equiv (a, b^2(a - c))$, from which it is seen that the tangent at T passes through P.

When the curve has a maximum and a minimum, if the minimum is above the X-axis, a is greater than c and the ordinate of T is positive. Hence $b^2(a-c)$ is positive, and b is therefore real; but if the finite maximum is below the X-axis, a is less than c and the ordinate of T is negative. Hence $b^2(a-c)$ is negative and, since $a-c$ is negative, b is real. In the third case the X-axis is cut in three places and we may choose any of the three intersections for c. If the intersection corresponding to the root whose value is algebraically least is chosen, and a is greater than c, then $b^2(a-c)$ is negative; but since $a-c$ is positive, b is a pure imaginary. If the middle intersection is taken for c, $b^2(a-c)$ is either positive or negative and a is greater or less than c, in such wise that b is always a pure imaginary. If the middle intersection, when it is taken for c, is at a point of inflection, $b^2(a-c)=0$ and $a=c$. In this case b cannot be determined graphically. If the third intersection is taken for c, $b^2(a-c)$ is positive and a is less than c, and b is therefore a pure imaginary. Hence in the third case $a \pm b\sqrt{-1}$ is real. It is seen that the third case furnishes an interesting check in graphically obtaining the roots of the cubic when they are all real. If the curve possesses no maximum and minimum we may substitute *point of inflection* for both maximum and minimum in the foregoing discussion.

PASADENA, CAL.,
APRIL, 1908.

THE TOPOGRAPHY OF CERTAIN CURVES DEFINED BY A DIFFERENTIAL EQUATION

By F. R. Sharpe

Consider the equation

$$(1)\qquad \frac{dy}{dx} = \frac{c_1}{c_2} = \frac{ax^2 + 2hxy + by^2 + 2gx + 2fy + c}{a'x^2 + 2h'xy + b'y^2 + 2g'x + 2f'y + c'}.$$

The existence and form of the solution near an ordinary point and near a singular point (that is a point of intersection of $c_1 = 0$, $c_2 = 0$) are well known. The object of this paper is to discuss the general configuration of the integral curves of (1).

The pencil of conics

$$(2)\qquad \frac{c_1}{c_2} = \lambda$$

through the four points of intersection of the conics $c_1 = 0$, $c_2 = 0$ is such that the integral curves of (1) have the same slope λ at the points where they cut the curves of (2). Massau* has given to such curves (2) the name *isoclines*. The slope of (2) is given by

$$(3)\qquad \frac{\partial c_1}{\partial x} + \frac{\partial c_1}{\partial y}\frac{dy}{dx} = \lambda\left(\frac{\partial c_2}{\partial x} + \frac{\partial c_2}{\partial y}\frac{dy}{dx}\right).$$

Hence (2) will have the slope λ at the two points in which it meets the line

$$(4)\qquad \frac{\partial c_1}{\partial x} + \lambda\frac{\partial c_1}{\partial y} = \lambda\frac{\partial c_2}{\partial x} + \lambda^2\frac{\partial c_2}{\partial y}.$$

Eliminating λ between (2) and (4) we find for the locus of all such points for the pencil of conics (2) the quintic

$$(5)\qquad \frac{\partial c_1}{\partial x}c_2^2 + \frac{\partial c_1}{\partial y}c_1c_2 = \frac{\partial c_2}{\partial x}c_1c_2 + \frac{\partial c_2}{\partial y}c_1^2.$$

If we differentiate (1) we see that (5) is also the locus of points on the system of integral curves of (1) at which

$$\frac{d^2y}{dx^2} = 0.$$

---

* D'Ocagne, *Calcul graphique*, p. 149.

It is therefore the locus of the points of inflection of the system of integral curves of (1).*

From the form of (5) it is easily seen that it represents a quintic having a double point at each of the four points of intersection of $c_1 = 0$, $c_2 = 0$. Each conic of the pencil meets the quintic in two points apart from the double points. The pencil of conics therefore determines on the quintic a group of two points with one degree of freedom. The tangents to (2) at these two points have the slope λ and are therefore parallel, the line (4) joining them must then be a diameter of the conic. The envelope of these lines is the conic†

$$\left(\frac{\partial c_1}{\partial y} - \frac{\partial c_2}{\partial x}\right)^2 + 4\frac{\partial c_1}{\partial x}\frac{\partial c_2}{\partial y} = 0.$$

There are three values of λ for which (2) breaks up into two straight lines. The quintic (5) therefore passes through the three points of intersection of the pairs of lines and (4) is in each case the tangent to the quintic at the intersection. If one of the lines into which (2) breaks up has the slope λ the quintic degenerates into this line and a quartic. If (2) breaks up into two parallel lines, (4) is an asymptote of the quintic. There are two values of λ for which (2) is a parabola. The line (4) is then parallel to the axis of the parabola and one point of the quintic is at infinity. Hence (4) is parallel to an asymptote of the quintic. If the axis of the parabola has the slope λ, the quintic touches the line at infinity.

If we choose for our conics two other conics of the pencil, the differential equation (1) becomes

$$\frac{dy}{dx} = \frac{Ac_1 + Bc_2}{Cc_1 + Dc_2}. \tag{6}$$

The corresponding quintics therefore form a triply infinite system each having a double point at the four points of intersection of the pencil of conics which determines twelve of the twenty constants of the general quintic. The quintics also pass through the three intersections of the pairs of lines of the pencil and through the two points at infinity on the axes of the two parabolas of the pencil. Hence we have accounted for all the constants.

The asymptotes of (2) are parallel to the lines

$$ax^2 + 2hxy + by^2 = \lambda(a'x^2 + 2h'xy + b'y^2).$$

* D'Ocagne, *Calcul graphique*, p. 153.

† The envelope of the line joining a g_1^2 on any curve is a conic. See Bobek, *Wiener Berichte*, vol. 93, part 2 (1886).

Hence there are three values of λ for which an asymptote of (2) is also an asymptote of the quintic, namely the three roots of

$$a + 2\lambda + b\lambda^2 = \lambda(a' + 2h'\lambda + b'\lambda^2),$$

and (4) is the equation of the asymptote.

The quintic corresponding to (6) is therefore determined. When the conic (2) is an ellipse or parabola the two points on it which also lie on (5) are necessarily real, but if (2) is a hyperbola the two points may be imaginary so that no part of the quintic lies in one pair of the angles formed by two of its asymptotes.

When (2) has the slope λ at one of the intersections of $c_1 = 0$, $c_2 = 0$, it is easily seen geometrically that the tangent to (2) is also one of the two tangents to the quintic at the intersection. On moving the origin to the intersection, (1) takes the form

$$(7) \qquad \frac{dy}{dx} = \frac{ax^2 + 2hxy + by^2 + 2g_1x + 2f_1y}{a'x^2 + 2h'xy + 2b'y^2 + 2g_1'x + 2f_1'y}.$$

The tangent at the origin to (2) is therefore

$$g_1x + f_1y = \lambda(g_1'x + f_1'y),$$

and if this has the slope λ,

$$g_1 + f_1\lambda = \lambda(g_1' + f_1'\lambda).$$

Hence the point is a crunode, cusp, or acnode of the quintic according as the roots of this quadratic are real, equal, or imaginary. The type of singularity at the origin depends, as is well known, on the nature of the roots of this quadratic. When the origin is an acnode the integral curves near the origin must be spirals since there is no inflection near the origin and the integral curves cut each conic of the pencil at an angle different from zero, since they can be tangent only at points of the quintic.

When two or more of the four points of intersection of $c_1 = 0, c_2 = 0$ coincide, the pencil of conics have a common tangent at the point and if this point be taken for origin, (1) has the form

$$(8) \qquad \frac{dy}{dx} = \frac{m(2gx + 2fy) + ax^2 + 2hxy + by^2}{m'(2gx + 2fy) + a'x^2 + 2h'xy + b'y^2}.$$

The slope of the integral curve through a point (x, y), which approaches the origin along a curve not there tangent to $gx + fy = 0$, approaches the

value m/m'. By approaching the origin along one of the conics of the pencil, dy/dx may be made to take any desired value λ. On making the linear substitution

$$x' = mx - m'y,$$
$$y' = gx + fy,$$

the equation (8) takes the form, dropping the accents,

$$\frac{dy}{dx} = \frac{2y + ax^2 + 2hxy + by^2}{a'x^2 + 2h'xy + b'y^2}, \tag{9}$$

in which a, a', etc. do not however denote the same values as in (8). This substitution fails in the case $\frac{m}{m'} = -\frac{g}{f}$, that is when the slope of the integral curves is the same as the slope of the common tangent of the pencil of conics. In this case we take

$$x' = x,$$
$$y' = gx + fy,$$

and (8) takes the form

$$\frac{dy}{dx} = \frac{a'x^2 + 2h'xy + b'y^2}{2y + ax^2 + 2hxy + by^2}. \tag{10}$$

The integral curves of the two normal forms (9) and (10) are of the same nature as orthogonal trajectories, the values of dy/dx from (9) and (10) being reciprocal. If three of the four intersections of the pencil of conics coincide, $a' = 0$, and the conics have second order contact. If all the intersections coincide, $a' = h' = 0$, and the conics have third order contact. If $a'x^2 + 2h'xy + b'y^2$ is a perfect square, the conics have double contact, and $\sqrt{a'}\,x + \sqrt{h'}\,y = 0$ is the common chord.

The quintic corresponding to (9) is, from (5),

$$\begin{aligned}&(ax + hy)(a'x^2 + 2h'xy + b'y^2)^2\\ &\quad + (1 + hx + by - a'x - h'y)(a'x^2 + 2h'xy + b'y^2)(2y + ax^2 + 2hxy + by^2)\\ &\quad - (h'x + b'y)(2y + ax^2 + 2hxy + by^2)^2 = 0.\end{aligned}$$

The terms of lowest degree are $2y(a'x^2 - b'y^2)$ and the terms which are independent of y are $(ha' - h'a)ax^5 + ca'x_4$. Hence the origin is a triple point, $y = 0$ being a tangent, and y is a factor if $a' = h' = 0$, or when $a = 0$.

When $a' = 0$ there is a cusp at the origin as well as a simple branch. When $a'x^2 + 2h'xy + b'y^2$ is a perfect square, $\sqrt{a'}\,x + \sqrt{b'}\,y$ is a factor. If $a' = h' = 0$, the quartic to which the quintic reduces has a tacnode as the origin of the form

$$2y + ax^2 = 0, \quad y + ax^2 = 0.$$

The quintic corresponding to (10) is

$$(a'x + h'y)(2y + ax^2 + 2hxy + by^2)^2$$
$$+ (\overline{h' - a}\,x + \overline{b' - h}\,y)(a'x^2 + 2h'xy + b'y^2)\ (2y + ax^2 + 2hxy + by^2)$$
$$- (1 + hx + by)(a'x^2 + 2h'xy + b'y^2)^2 = 0.$$

The terms of lowest degree are $4(a'x + h'y)y^2$, and the terms which are independent of y are $(h'a - ha')a'x^5 - a'^2x^4$. Hence the origin is a triple point with a cusp. When $a' = 0$ then y is a factor, and when $a' = h' = 0$ the quintic reduces to a cubic.

In any particular example it is usually easy to draw certain conics of the pencil and to mark on the inflectional quintic the values of λ from $-\infty$ to $+\infty$. The integral curves are then seen to be of certain distinct types depending on the nature of the four singular points. The following are interesting cases.

EXAMPLE 1.

$$\frac{dy}{dx} = \frac{(x + y)^2 - 1}{(x - y)^2 - 1}.$$

Here $(x + y)^2 = 1$ and $(x - y)^2 = 1$ are the lines and degenerate parabolas of the pencil. Hence $x + y = 0$ and $x - y = 0$ are asymptotes of the quintic. The points $(\pm 1, 0)$ are acnodes and the points $(0, \pm 1)$ crunodes, the tangents having slopes $1 \pm \sqrt{2}$. An asymptote of the conic of the pencil $\lambda = 3.38$ is the asymptote $y = 3.38x$ of the quintic.

The quintic has its center at the origin (see the figure).

EXAMPLE 2.

$$\frac{dy}{dx} = \frac{-x - xy}{4x + 4y + 4xy}.$$

The quintic in this case reduces to

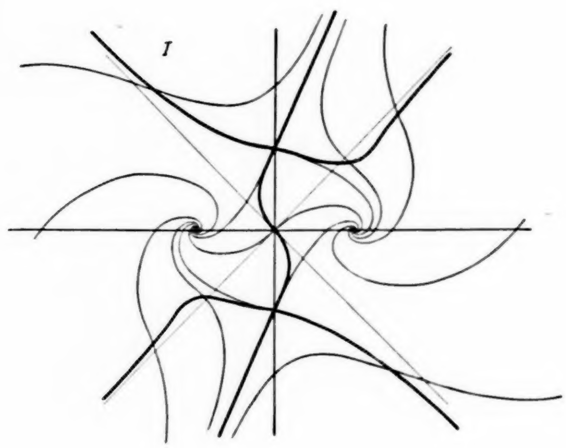
$$(y + 1)(x^2 + 4xy + 4y^2 + 4xy^2) = 0.$$

Example 3.

$$\frac{dy}{dx} = \frac{y^2 - 1}{x^2}, \quad \text{and} \quad \frac{dy}{dx} = \frac{-x^2}{y^2 - 1}.$$

The quintics reduce to

$$x^2(y^2 - 1)(y - x) = 0, \; x(y^2 - 1)^2 = x^4 y.$$



The integral curves are

$$y = \frac{1 + Ce^{-2x}}{1 - Ce^{-2x}}, \quad \frac{y^3}{3} - y + \frac{x^3}{3} = C.$$

For the first of these we have $\lim\limits_{x=0+} (y) = 1$, and $\lim\limits_{x=0-} (y) = -1$.

Example 4.

$$\frac{dy}{dx} = \frac{y - x^2}{y^2}.$$

The quintic reduces to $y = 0$, and $(y - x^2)(y - 2x^2) + 2xy^3 = 0$.

Cornell University
Ithaca, N. Y.,
February, 1909.

ABEL'S THEOREM AND THE ADDITION FORMULAE FOR ELLIPTIC INTEGRALS

By Harry Huntington Barnum

The addition formulae for the Legendrian elliptic integrals

$$1a) \qquad F(k, z) = \int_0^z \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

$$1b) \qquad E(k, z) = \int_0^z \frac{(1-k^2z^2)\,dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

$$1c) \qquad \Pi(n, k, z) = \int_0^z \frac{dz}{(1+nz^2)\sqrt{(1-z^2)(1-k^2z^2)}},$$

are well known and have been derived in various ways. The object of the formulae is to give values for the expressions

$$F(k, z_1) + F(k, z_2),$$
$$E(k, z_1) + E(k, z_2),$$
$$\Pi(n, k, z_1) + \Pi(n, k, z_2),$$

each in terms of a third integral of the same type, whose upper limit is an algebraic function of z_1 and z_2, with the addition of an algebraic or a logarithmic term in the second and third cases.

In most of the elementary works on the subject the proofs of the formulae depend on the more or less artificial discovery of an algebraic solution of the differential equation

$$\frac{dz_1}{\sqrt{(1-z_1^2)(1-k^2z_1^2)}} + \frac{dz_2}{\sqrt{(1-z_2^2)(1-k^2z_2^2)}} = 0,$$

or upon somewhat complicated applications of Abel's Theorem. The object of this paper is to give an elementary derivation of the addition formulae

suitable for presentation to a beginning class in the subject, by a method which may serve also as an introduction to Abel's famous theorem.

A complete comprehension of Abel's theorem as it is about to be stated* requires an acquaintance with the theory of functions of a complex variable, and in particular with Riemann surfaces. But the succeeding sections are developed independently of the theorem, its use being merely to indicate the methods to be used. Furthermore the lower limits of the integrals involved in the proofs are everywhere so arranged that only real expressions occur.

For the sake of brevity the addition formula for integrals of the first of the three types (1) is derived from that of the third type by putting $n = 0$. But the simple independent proof, quite similar to that of the third type can be readily carried through by the reader.

1. Abel's Theorem. Let C and C' be plane curves given by the equations

$$(2) \qquad \begin{aligned} C &: F(xy) = 0, \\ C' &: \phi(xy) = 0. \end{aligned}$$

These curves have n points of intersection $(x_1, y_1), \cdots (x_n, y_n)$, where n is the product of the degrees of C and C'. Let $R(x,y)$ be a rational function of x and y where y is defined as a function of x by the relation $F(x,y) = 0$.

Consider the sum

$$(3) \qquad I = \sum_{i=1}^{n} \int_{x_0 y_0}^{x_i y_i} R(x,y)\,dx,$$

the integrals $\int_{x_0 y_0}^{x_i y_i} R(xy)\,dx$ being taken from a fixed point, (x_0, y_0) in the Riemann surface of the function $F(x,y)$, to the n points of intersection $(x_1, y_1), \cdots (x_n, y_n)$ of the curves C and C'. If some of the coefficients $a_1, a_2, \cdots a_k$ of $\phi(x,y)$ are regarded as continuous variables, the points (x_i, y_i) will vary continuously and hence I will be a function, whose form is to be determined, of the variable coefficients $a_1, a_2, \cdots a_k$.

Abel's theorem may now be stated as follows:

The partial derivative of the sum I, with respect to any one of the coefficients of the variable curve $\phi = 0$, is a rational function of the coefficients and hence I is equal to a rational function of the coefficients of

---

* See Goursat, *Cours d'Analyse*, vol. 2, §360.

$\phi(x,y) = 0$, plus a finite number of logarithms or arc tangents of such rational functions.

2. A useful lemma. Before proceeding to the derivation of the addition formulae a lemma* will be proved which will be of use later.

Let $f(x) = 0$ be an equation of the nth degree with the distinct roots $x_1, x_2, \cdots x_n$. Also let $F(x) = 0$ be any abitrary polynomial in x. The following identity can then be readily established,

$$\frac{xF(x)}{f(x)} = k + k_1 x + k_2 x^2 + \cdots + \sum_{i=1}^{n} \frac{x_i F(x_i)}{(x - x_i) f'(x_i)}. \tag{4}$$

The integral terms are the quotient of $xF(x)$ divided by $f(x)$, and the terms of the sum are the partial fractions to which the remainder over $f(x)$ gives rise. If we set $x = 0$ we get

$$\sum_{i=1}^{n} \frac{F(x_i)}{f'(x_i)} = k \tag{5}$$

from (4). Hence we see that *if* $F(x)$ *is of degree* $(n - 2)$ *or less, then* $k = 0$, *and the symmetric functions* (5) *of the roots must be zero. If* $F(x)$ *is of degree* $(n - 1)$ *or more*

$$\sum_{i=1}^{n} \frac{F(x_i)}{f'(x_i)}$$

will be equal to the term independent of x *in the quotient* $xF(x)/f(x)$, as is indicated by (5).

3. The Addition Formula for $E(k, z)$. We proceed now to the establishment of the addition formula for elliptic integrals of the second type,

$$E(k, z) = \int_0^z \frac{(1 - k^2 z^2)\, dz}{\sqrt{(1 - z^2)(1 - k^2 z^2)}}. \tag{6}$$

By the substitution $z^2 = x$ (6) may be thrown into the more convenient form

$$E(k, x) = \int_0^x \frac{(1 - k^2 x)\, dx}{\sqrt{x(1 - x)(1 - k^2 x)}},$$

* Petersen: *Funktionstheorie*, §39.

which is the Riemann normal form. We take now, corresponding to the curves C and C' of (2), the curves

$$(7a) \qquad C: \qquad y^2 = x(1-x)(1-k^2x),$$

$$(7b) \qquad C': \qquad y = ax + b.$$

The elimination of y between these two equations will give us as the abscissae x_1, x_2, x_3 of the points of intersection the three roots of the equation

$$(8) \qquad \phi(x) = k^2x^3 - (1 + k^2 + a^2)\, x^2 + (1 - 2ab)x - b^2 = 0.$$

The function $R(x, y)$ in (3) is here

$$R(x, y) = \frac{1 - k^2x}{y},$$

where y is defined by (7a). We have then corresponding to the sum (3), the function

$$(9) \qquad I(a, b) = \int_0^{x_1} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_{\frac{1}{k^2}}^{x_3} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}}.$$

For the purpose of keeping the discussion in the real field the lower limit of the third integral is taken as $1/k^2$ instead of 0. As however the integral between the limits 0 and $1/k^2$ is a constant, this is not an essential alteration of the sum which occurs in Abel's theorem.

To determine the form of $I(a, b)$ we take its partial derivatives with respect to a and b

$$(10) \qquad \begin{aligned} \frac{\partial I}{\partial a} &= \frac{\partial I}{\partial x_1}\frac{\partial x_1}{\partial a} + \frac{\partial I}{\partial x_2}\frac{\partial x_2}{\partial a} + \frac{\partial I}{\partial x_3}\frac{\partial x_3}{\partial a} \\ &= \frac{1-k^2x_1}{y_1}\frac{\partial x_1}{\partial a} + \frac{1-k^2x_2}{y_2}\frac{\partial x_2}{\partial a} + \frac{1-k^2x_3}{y_3}\frac{\partial x_3}{\partial a}. \end{aligned}$$

Since x_1 is a root of (8) we may substitute x_1 for x in it and take the derivative with respect to a, noting that x_1 is a function of a. We find then

$$\phi'(x_1)\frac{\partial x_1}{\partial a} + \frac{\partial \phi(x_1)}{\partial a} = 0,$$

the accent denoting differentiation with respect to x_1. From equation (8) this becomes

$$\phi'(x_1)\,\frac{\partial x_1}{\partial a} - 2(ax_1 + b)x_1 = 0,$$

or because of (7b)

$$\phi'(x_1)\,\frac{\partial x_1}{\partial a} - 2y_1x_1 = 0.$$

Substituting this value for $\frac{\partial x_1}{\partial a}$ and the similar ones for $\frac{\partial x_2}{\partial a}$ and $\frac{\partial x_3}{\partial a}$ in (10), we get

$$\frac{\partial I}{\partial a} = 2\sum_{i=1}^{3}\frac{x_i}{\phi'(x_i)} - 2k^2\sum_{i=1}^{3}\frac{x_i^2}{\phi'(x_i)}.$$

These sums are in the form of those in (5), where $f(x)$ is the function $\phi(x)$ in equation (8), and for the first sum $F(x) = x$, while for the second $F(x) = x^2$. Hence because of the lemma the first sum vanishes and the second is equal to the constant term $1/k^2$ of the quotient $x^3/\phi(x)$. We have therefore,

$$\frac{\partial I}{\partial a} = -\,2. \tag{11}$$

Similarly

$$\frac{\partial I}{\partial b} = \frac{\partial I}{\partial x_1}\frac{\partial x_1}{\partial b} + \frac{\partial I}{\partial x_2}\frac{\partial x_2}{\partial b} + \frac{\partial I}{\partial x_3}\frac{\partial x_3}{\partial b}$$

$$= 2\sum_{i=1}^{3}\frac{1}{\phi'(x_i)} - 2k^2\sum_{i=1}^{3}\frac{x_i}{\phi'(x_i)} = 0. \tag{12}$$

To find $I(a, b)$ we now integrate equations (11) and (12). Equation (11) gives

$$I(a, b) = -\,2a + f(b).$$

The derivative of this expression with respect to b vanishes because of (12), and hence $f(b)$ must be a constant. Consequently equation (9) becomes

$$\int_0^{x_1}\frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2}\frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_{\frac{1}{k^2}}^{x_3}\frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} = -\,2a + C, \tag{13}$$

where x_1, x_2, x_3 are the abscissae of the intersections of the curve (7*a*) with the straight line (7*b*), i. e., the three roots of equation (8). The formula (13) is in fact an addition formula for $E(k, x)$, though not in the usual form.

In order to transform equation (13) let us keep the intersection point (x_3, y_3) in figure 1 fixed, place (x_1, y_1) at the origin, and denote the resulting

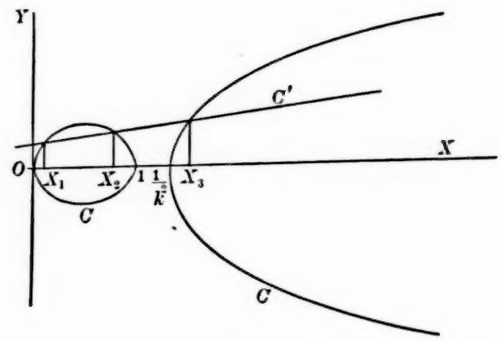


FIG. 1.

values of x_2, y_2 and a by x', y' and a' respectively. Then equation (13) becomes

$$\int_{\frac{1}{k^2}}^{x_3} \frac{(1-k^2x)dx}{\sqrt{x(1-x)(1-k^2x)}} = -\int_0^{x'} \frac{(1-k^2x)dx}{\sqrt{x(1-x)(1-k^2x)}} - 2a' + C,$$

which substituted in (13) gives

$$\int_0^{x_1} \frac{(1-k^2x)dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{(1-k^2x)dx}{\sqrt{x(1-x)(1-k^2x)}}$$
$$= \int_0^{x'} \frac{(1-k^2x)dx}{\sqrt{x(1-x)(1-k^2x)}} + 2(a'-a),$$

the usual form for the addition theorem.

It remains only to express x' and $a' - a$ in terms of x_1 and x_2. Since x_1, x_2, x_3 are the roots of equation (8), we have

$$(14) \quad x_1x_2x_3 = \frac{b^2}{k^2}, \quad x_2x_3 + x_3x_1 + x_1x_2 = \frac{1-2ab}{k^2}, \; x_1 + x_2 + x_3 = \frac{1+k^2+a^2}{k^2},$$

from which we find, when $x_1 = 0$,

$$b = 0, \quad x' = \frac{1}{k^2 x_3}, \quad a' = \sqrt{\frac{(1-x')(1-k^2x')}{x'}}. \tag{15}$$

From equations ($7b$) and (14)

$$\begin{aligned} a = \frac{y_3}{x_3} - \frac{b}{x_3} &= \sqrt{\frac{(1-x_3)(1-k^2x_3)}{x_3}} - k\sqrt{\frac{x_1x_2}{x_3}} \\ &= \sqrt{\frac{(1-x')(1-k^2x')}{x'}} - k^2\sqrt{x_1x_2x'} \\ &= a' - k^2\sqrt{x_1\, x_2\, x'}. \end{aligned} \tag{16}$$

From equations (14) and (15)

$$x' = \frac{1}{k^2x_3} = \frac{x_1x_2}{b^2}.$$

Since (x_1, y_1) and (x_2, y_2) are on the straight line ($7b$), the value of b can be found from the equations

$$y_1 = ax_1 + b, \quad y_2 = ax_2 + b.$$

We have then

$$b = \frac{x_1y_2 - x_2y_1}{x_1 - x_2} = x_1\, x_2\, \frac{1 - k^2x_1x_2}{x_1y_2 + x_2y_1},$$

where the second form is found by rationalizing the numerator of the first by the help of equation ($7a$), which (x_1,y_1) and (x_2,y_2) also satisfy. The final expression for x' is therefore

$$x' = \frac{1}{x_1x_2}\left(\frac{x_1y_2 + x_2y_1}{1 - k^2x_1x_2}\right)^2. \tag{17}$$

The addition formula for Elliptic integrals of the second type in the Riemann normal form is

$$\begin{aligned} \int_0^{x_1} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} \\ = \int_0^{x'} \frac{(1-k^2x)\,dx}{\sqrt{x(1-x)(1-k^2x)}} + 2k^2\sqrt{x_1x_2x'}, \end{aligned} \tag{18}$$

where x' is defined in terms of x_1 and x_2 by equation (17). The addition formula for the Legendre normal form can be found by transforming the variables in (17) and (18) by the substitution $x = z^2$.

It follows that *the addition formula for $E(k, z)$ is*

$$(19)\qquad \int_0^{z_1} \frac{(1-k^2z^2)dz}{\sqrt{(1-z^2)(1-k^2z^2)}} + \int_0^{z_2} \frac{(1-k^2z^2)dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = \int_0^{z} \frac{(1-k^2z^2)dz}{\sqrt{(1-z^2)(1-k^2z^2)}} + k^2z_1z_2z',$$

where z' is expressible in the form

$$(20)\qquad z' = \frac{z_1\sqrt{(1-z_2^2)(1-k^2z_2^2)} + z_2\sqrt{(1-z_1^2)(1-k^2z_1^2)}}{1-k^2z_1^2z_2^2}.$$

The addition formula for the usual trigonometric form of $E(k, z)$ can be readily found by making the substitution $z = \sin\phi$ on all the variables in (19) and (20).

4. The addition formula for $\Pi(n, k, z)$. We proceed now to the treatment of elliptic integrals of the third type, with the help again of the Riemann normal form. We take as the curves C and C' in (2) the same curves (7), the function $R(x, y)$ being in this case, however,

$$R(x, y) = \frac{1}{(1+nx)y}.$$

The sum analagous to that occurring in Abel's theorem is here

$$(21)\qquad K(ab) = \int_0^{x_1} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}} + \int_{\frac{1}{k^2}}^{x_3} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}}.$$

The partial derivatives of K with respect to a and b are

$$(22)\qquad \frac{\partial K}{\partial a} = \frac{\partial K}{\partial x_1}\frac{\partial x_1}{\partial a} + \frac{\partial K}{\partial x_2}\frac{\partial x_2}{\partial a} + \frac{\partial K}{\partial x_3}\frac{\partial x_3}{\partial a} = 2\sum_{i=1}^{3} \frac{x_i}{(1+nx_i)\phi'(x_i)}.$$

In order to apply the lemma in (5) we take

$$F(x) = x,$$
$$f(x) = k^2(1 + nx)(x - x_1)(x - x_2)(x - x_3),$$

with the roots $-1/n$, x_1, x_2, x_3. We have then

$$\sum_{i=1}^{n} \frac{F(x_i)}{f'(x_i)} = \frac{\frac{1}{n}}{k^2 n\left(\frac{1}{n} + x_1\right)\left(\frac{1}{n} + x_2\right)\left(\frac{1}{n} + x_3\right)} + \sum_{i=1}^{3} \frac{x_i}{(1 + nx_i)\phi'(x_i)}.$$

According to the lemma this sum must be zero, which gives us

$$\sum_{i=1}^{3} \frac{x_i}{(1 + nx_i)\phi'(x_i)} = \frac{-n}{k^2(1 + nx_1)(1 + nx_2)(1 + nx_3)}.$$

The substitution of this value in (22) gives

$$\frac{\partial K}{\partial a} = \frac{-2n}{k^2(1 + nx_1)(1 + nx_2)(1 + nx_3)}.$$

This may be written in the form

$$\frac{\partial K}{\partial a} = \frac{-2n}{k^2[1 + n(x_1 + x_2 + x_3) + n^2(x_2x_3 + x_3x_1 + x_1x_2) + n^3x_1x_2x_3]}.$$

The application of the relations given in (14) reduces this to

$$\frac{\partial K}{\partial a} = \frac{-2}{a^2 - 2nba + 1 + k^2 + n + nb^2 + \frac{k^2}{n}}. \tag{23}$$

In a similar manner, by using $F(x) = 1$ in the lemma, the value of $\frac{\partial K}{\partial b}$ is found to be

$$\frac{\partial K}{\partial b} = \frac{2n}{n^2b^2 - 2anb + n + 1 + k^2 + a^2 + \frac{k^2}{n}}. \tag{24}$$

The integration of equation (23) can be effected by the use of a table of integrals. If we confine ourselves to real variables, there are two cases ac-

cording as $p = (1 + n)(1 + k^2/n)$ is greater than or less than zero. In the former case, we have

$$(25) \qquad p > 0, \quad K = \frac{2}{\sqrt{p}} \text{ arc tan } \frac{nb - a}{\sqrt{p}} + f_1(b).$$

For the latter case,

$$(26) \qquad p < 0, \quad K = \frac{1}{\sqrt{-p}} \log \frac{nb - a - \sqrt{-p}}{nb - a + \sqrt{-p}} + f_2(b).$$

By means of the formula

$$\text{arc tan } x = \frac{1}{2i} \log \frac{i + x}{i - x}, \qquad i = \sqrt{-1}$$

the two cases could be treated simultaneously. As it is our object however to keep the discussion in the real field the two cases will be treated separately.

Differentiation of equations (25) and (26) with respect to b, gives us

$$\frac{\partial K}{\partial b} = \frac{2n}{n^2b^2 - 2anb + n + 1 + k^2 + a^2 + \dfrac{k^2}{n}} + f'(b),$$

and comparison with equation (24) shows us that $f'(b) = 0$, and hence $f(b) = C$ in each case.

From this point the work will be carried through only for the function (25), since the method is the same for each. Equation (21) then becomes

$$(27) \qquad \begin{aligned} &\int_0^{x_1} \frac{dx}{(1 + nx)\sqrt{x(1 - x)(1 - k^2x)}} + \int_0^{x_2} \frac{dx}{(1 + nx)\sqrt{x(1 - x)(1 - k^2x)}} \\ &\quad + \int_{\frac{1}{k^2}}^{x_3} \frac{dx}{(1 + nx)\sqrt{x(1 - x)(1 - k^2x)}} = \frac{2}{\sqrt{p}} \text{ arc tan } \frac{nb - a}{\sqrt{p}} + C. \end{aligned}$$

Placing the point (x_1, y_1) at the origin, and using the notation of §3, we find again a value for the third integral in this expression:

$$\int_{\frac{1}{k^2}}^{x_3} \frac{dx}{(1 + nx)\sqrt{x(1 - x)(1 - k^2x)}} = -\int_0^{x'} \frac{dx}{(1 + nx)\sqrt{x(1 - x)(1 - k^2x)}} + \frac{2}{\sqrt{p}} \text{ arc tan } \frac{-a'}{\sqrt{p}} + C.$$

Therefore (27) becomes

$$\int^{x_1} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}}$$
$$= \int_0^{x'} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}} + \frac{2}{\sqrt{p}}\left[\text{arc tan}\,\frac{nb - a}{\sqrt{p}} + \text{arc tan}\,\frac{a'}{\sqrt{p}}\right].$$

The trigonometric formula

$$\text{arc tan}\,\alpha + \text{arc tan}\,\beta = \text{arc tan}\,\frac{\alpha + \beta}{1 - \alpha\beta}$$

reduces the term outside the integral signs to the following form:

$$\text{arc tan}\,\frac{nb - a}{\sqrt{p}} + \text{arc tan}\,\frac{a'}{\sqrt{p}} = \text{arc tan}\,\frac{\sqrt{p}(nb - a + a')}{p - a'(nb - a)}.$$

The substitution of the values of a, a', and b from (14), (15), and (16) and the division of both numerator and denominator by the common factor $(n + k^2x')$ gives as *the final form of the addition formula of Elliptic integrals of the third type in the Riemann normal form, when* $p > 0$:

$$\int_0^{x_1} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}}$$
$$= \int_0^{x'} \frac{dx}{(1 + nx)\sqrt{x(1-x)(1-k^2x)}}$$
$$+ \frac{2}{\sqrt{p}}\,\text{arc tan}\,\frac{\sqrt{p}\,n\sqrt{x_1x_2x'}}{1 + nx' - n\sqrt{x_1x_2(1-x')(1-k^2x')}},$$

where x' *is expressible in terms of* x_1 *and* x_2 *by equation* (17).

The substitution $x = z^2$ as before gives us the corresponding *addition formula for the Legendre form*:

$$\int_0^{z_1} \frac{dz}{(1 + nz^2)\sqrt{(1-z^2)(1-k^2z^2)}} + \int_0^{z_2} \frac{dz}{(1 + nz^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$
$$= \int_0^{z'} \frac{dz}{(1 + nz^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$
$$+ \frac{1}{\sqrt{p}}\,\text{arc tan}\,\frac{\sqrt{p}\,nz_1z_2z'}{1 + nz'^2 - nz_1z_2\sqrt{(1-z'^2)(1-k^2z'^2)}},$$

where the expression for z' in terms of z_1 and z_2 is given in (20).

A similar treatment of the case where $p < 0$ brings us to the forms:

$$\int_0^{x_1} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}}$$

$$= \int_0^{x'} \frac{dx}{(1+nx)\sqrt{x(1-x)(1-k^2x)}}$$

$$+ \frac{1}{\sqrt{-p}} \log \frac{n\sqrt{x_1x_2(1-x')(1-k^2x')} - (1+nx') - \sqrt{-p}\, n\sqrt{x_1x_2x'}}{n\sqrt{x_1x_2(1-x')(1-k^2x')} - (1+nx') + \sqrt{-p}\, n\sqrt{x_1x_2x'}},$$

and

$$\int_0^{z_1} \frac{dz}{(1+nz^2)\sqrt{(1-z^2)(1-k^2z^2)}} + \int_0^{z_2} \frac{dz}{(1+nz^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$

$$= \int_0^{z'} \frac{dz}{(1+nz^2)\sqrt{(1-z^2)(1-k^2z^2)}}$$

$$+ \frac{1}{2\sqrt{p}} \log \frac{nz_1z_2\sqrt{(1-z'^2)(1-k^2z'^2)} - (1+nz'^2) - \sqrt{-p}\, nz_1z_2z'}{nz_1z_2\sqrt{(1-z'^2)(1-k^2z'^2)} - (1+nz'^2) + \sqrt{-p}\, nz_1z_2z'}.$$

As stated at the beginning of the paper, the addition formula for Elliptic integrals of the first type may be found directly from those of the third type by putting $n = 0$. From (23) and (24) we see that if $n = 0$, then $K(ab)$ is a constant, and hence *the addition formula for Elliptic integrals of the first type reduces to the simple forms*:

$$\int_0^{x_1} \frac{dx}{\sqrt{x(1-x)(1-k^2x)}} + \int_0^{x_2} \frac{dx}{\sqrt{x(1-x)(1-k^2x)}} = \int_0^{x'} \frac{dx}{\sqrt{x(1-x)(1-k^2x)}},$$

and

$$\int_0^{z_1} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} + \int_0^{z_2} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}} = \int_0^{z'} \frac{dz}{\sqrt{(1-z^2)(1-k^2z^2)}},$$

where, as before, x' and z' are related to x_1, x_2 and z_1, z_2 respectively by equations (17) and (20).

PRINCETON, UNIVERSITY,
PRINCETON, N. J.

ON THE DETERMINATION OF THE ASYMPTOTIC DEVELOPMENTS OF A GIVEN FUNCTION

By Walter B. Ford

1. Introduction. The determination of the asymptotic developments of a given function is usually a problem of considerable difficulty and, when regarded in a general sense, is one for which but fragmentary results exist at the present time. The known determinations appear to be either those for special functions such as Bessel's function $J_n(z)$* or for certain types of integral functions defined either by infinite products or by Maclaurin series.† The importance of such developments is, however, well known. In particular, if $f(z)$ be an integral function of z defined by means of a Weierstrass product,‡ the point $z = \infty$ will in general be essentially singular and there will be no direct means of determining the behavior of $f(z)$ in the neighborhood of this point since the given product is not adapted in form to the study of the function when $|z|$ is large. Such determinations are often important and are supplied as soon as the asymptotic developments are known. Likewise, asymptotic developments in general supply information regarding the behavior of functions in the neighborhood of the point infinity.

The problem, when stated in a precise form and in the one which we shall understand throughout the present paper, may be described as follows: Let $F(z)$ be a given function of the complex variable z defined throughout the finite z plane and such that (a) the point $z = \infty$ is an essentially singular point; and (b) when $|z|$ is sufficiently large and $arg\ z$ lies within some specified sector Λ there exist two functions $f_\lambda(z)$ and $\phi_\lambda(z)$, defined for values of z in Λ and such that for the same values of z

$$(1)\quad F(z)=f_\lambda(z)+\phi_\lambda(z)\left[a_{0,\lambda}+\frac{a_{1,\lambda}}{z}+\frac{a_{2,\lambda}}{z^2}+\cdots+\frac{a_{n-1,\lambda}}{z^{n-1}}+\frac{a_{n,\lambda}+w_{\lambda,n}(z)}{z^n}\right],$$

* Cf. Lommel, *Studien über die Bessel'schen Functionen* § 17 (1868).

† Cf. Barnes, *Philosophical Transactions*, Vol. 199 A, pp. 411–500 (1902); also Vol. 206 A, pp. 249–297 (1906). Each of these memoirs contains an excellent bibliography of the subject. Cf. also Mattson, *Contributions à la Théorie des Fonctions entières.* (Thèse) Upsal, 1905.

‡ Cf. Picard, *Traité d'Analyse*, Vol. II, pp. 140-143.

where $a_{0,\lambda}$, $a_{1,\lambda}$, $\cdots$ $a_{n,\lambda}$ (n arbitrary) are constants and

$$\lim_{|z|=\infty} w_{\lambda,n}(z) = 0.^*$$

To determine the functions $f_\lambda(z)$, $\phi_\lambda(z)$ *and the constants* $a_{0,\lambda}$, $a_{1,\lambda}$, $\cdots$ $a_{n,\lambda}$.†

In the present paper it is proposed to show how the so-called Maclaurin sum-formula may be used in some cases to solve the above problem.‡ For this purpose we shall apply the formula to a variety of special functions $F(z)$. No attempt will be made to obtain theorems of great generality, the belief being that a few illustrations will suffice to enable the reader to apply the method wherever possible for himself. In each of the cases considered only the functions $f_\lambda(z)$, $\phi_\lambda(z)$ and the first one of the constants $a_{0,\lambda}$, $a_{1,\lambda}$, $\cdots$ $a_{n,\lambda}$ which is not equal to zero are determined since these three determinations constitute what is essential to the study of the behavior of the function for large values of $|z|$. The method, however, permits equally of the determination of $a_{n,\lambda}$ when $n = 0, 1, 2, 3, \cdots$.

2. The Maclaurin Sum-Formula. The form in which we shall take the sum-formula above referred to is as follows:§ Let $f(x)$ be a function of the real variable x which together with its first $2k$ derivatives ($k \geqq 1$) is finite and continuous throughout the interval $x \geqq 0$.‖ Also, let

---

* It may be remarked that a function $F(z)$ may frequently be written as a linear combination of expressions of the form (1) when $|z|$ is sufficiently large and *arg* z is properly chosen, as in the case of Bessel's function $J\ (z)$ (cf. Lommel l. c.). In such cases the resulting development is likewise said to be asymptotic, but in the present paper we shall confine ourselves to the primary form (1).

† Frequently the proof of the *existence* of (b) constitutes a separate preliminary problem. This is the case for the functions considered in the present paper where both the existence and determination of the developments (1) are considered.

‡ This formula likewise constitutes the central feature of the method employed by Barnes in the first of the memoirs cited above (l. c. p. 444). The form in which it there occurs is essentially different, however, from that which we adopt, the latter appearing preferable because it takes into account the *remainder*, thereby rendering possible a greater degree of rigor and clearness in the deductions; also because certain restrictions present in the former case may be dispensed with, reference being here made especially to those imposed upon $\phi(z)$ (l. c., §41).

§ Cf. Markoff, *Differenzenrechnung* (Leipzig, 1896), pp. 98, 99, 132. Throughout we shall take $h = 1$.

‖ No conditions are explicitly stated by Markoff as regards the function $f(x)$, but it appears from the analysis of pp. 112, 113, 114 that the conditions which we here impose are sufficient.

$$\phi_q(t) = \frac{t^q}{q!} + A_1 \frac{t^{q-1}}{(q-1)!} + A_2 \frac{t^{q-2}}{(q-2)!} + \cdots + A_{q-1}t,$$

where $$A_1 = -\tfrac{1}{2},\ A_3 = A_5 = A_7 = \cdots = A_{2q+1} = 0$$

and $$A_{2q} = \frac{(-1)^{q-1}B_q}{(2q)!},$$

B_q representing the qth Bernoulli number.

Then, whenever the series

$$(2) \qquad \Omega_k(m) = \sum_{n=1}^{n=\infty} \int_0^1 f^{(2k)}(m+n-t)\phi_{2k}(t)\,dt \qquad (m \geqq a)$$

converges we may write

$$(3) \quad \sum_{x=a}^{x=m-1} f(x) = C_k + \int_a^m f(x)\,dx + A_1 f(m) + A_2 f'(m) + A_4 f'''(m) + \cdots + $$
$$+ A_{2k-2} f^{(2k-3)}(m) + \Omega_k(m), \qquad (a \geqq 0)$$

where C_k is a constant (independent of m) defined by the equation

$$C_k = -[A_1 f(a) + A_2 f'(a) + A_4 f'''(a) + \cdots + A_{2k-2} f^{(2k-3)}(a) + \Omega_k(a)].$$

For the important case in which $k = 1$ the formula becomes

$$(4) \qquad \sum_{x=a}^{x=m-1} f(x) = C + \int_a^m f(x)\,dx - \tfrac{1}{2}f(m) + \Omega_1(m),$$

where $$C = \tfrac{1}{2}f(a) - \Omega_1(a).$$

3. Application to the Study of Asymptotic Developments.

Example 1. *To obtain asymptotic developments for the function*

$$F(z) = \sum_{n=0}^{n=\infty} \frac{1}{(2n+1)^2 + z^2}.$$

We here choose a function which, by virtue of the well known formula

$$\frac{\tan z}{2z} = \sum_{n=0}^{n=\infty} \frac{1}{\frac{(2n+1)^2}{4}\pi^2 - z^2}$$

may readily be evaluated in the form

$$F(z)=\frac{\pi}{4z}\,\frac{e^{\frac{\pi z}{2}}-e^{-\frac{\pi z}{2}}}{e^{\frac{\pi z}{2}}+e^{-\frac{\pi z}{2}}},$$

thus enabling us to compare the results obtained with known facts.

Considering at first that z is real but different from zero, let us place

$$f(x)=\frac{1}{(2x+1)^2+z^2}, \qquad k=1,\ a=0.$$

Then the series (2) is convergent. In fact, upon observing that $\phi_2(t)$ is negative for all values of t between $t=0$ and $t=1$, we may apply the first law of the mean for integrals and write the nth term of the series—viz.:

$$\int_0^1 f^{(2)}(m+n-t)\phi_2(t)\,dt,$$

where the indicated second derivative is with respect to m – in the form

$$f^{(2)}(m+n-\theta)\int_0^1 \phi_2(t)\,dt=-\tfrac{1}{12}f^{(2)}(m+n-\theta) \qquad (0<\theta<1)$$

from which the indicated result readily appears.

Likewise it appears that the nth term of the series (2) approaches the limit zero when $m=\infty$ and hence that $\lim\limits_{m=\infty}\Omega_1(m)=0$.

Upon employing formula (4) and placing $m=\infty$; also noting that $\lim\limits_{m=\infty} f(m)=0$, we thus obtain

$$F(z)=\int_0^1\frac{dx}{(2x+1)^2+z^2}+\frac{1}{2(1+z^2)}+R(z)$$

where

$$(5) \qquad R(z)=-\Omega_1(0)=-\sum_{n=1}^{n=\infty}\int_0^1\left\{\frac{d^2}{dx^2}\left[\frac{1}{(2x+1)^2+z^2}\right]\right\}_{x=n-t}\phi_2(t)\,dt.$$

We note also that since each term of the series (5) when multiplied by z vanishes when $z=\pm\infty$, we shall have $\lim\limits_{z=\pm\infty} zR(z)=0$.

Moreover, if in particular z is *positive* we may write

$$\int_0^1 \frac{dx}{(2x+1)^2+z^2} = \frac{1}{2z}\left[\text{arc tan}\,\frac{2x+1}{z}\right]_{x=0}^{x=\infty} = \frac{\pi}{4z} - \frac{1}{2z}\,\text{arc tan}\,\frac{1}{z},$$

so that for large positive values of z we obtain

$$(6) \qquad F(z) = \frac{\pi}{4z} + \frac{\epsilon(z)}{z}\;; \qquad \lim_{z=+\infty} \epsilon(z) = 0.$$

This last result may now be generalized to all values of z, real or complex, lying in any sector Λ whose vertex is at the point $z = 0$ and whose bounding lines lie within the right half of the z plane, the neighborhood of the point $z = 0$ being always understood to be excluded. To see this we first note that the series appearing in (5) not only converges, as we have indicated, for values of z which are real and different from zero, but for all values of z in Λ the convergence is likewise seen to exist and to be *uniform*. Moreover, each term of $R(z)$ is analytic throughout Λ. Thus it follows that $R(z)$ is itself analytic in this region.* Whence, by taking Λ so large that it includes the portion of the positive real axis in which (6) holds, the two members of the same equation come to represent two functions of the complex variable z, each analytic throughout Λ and coinciding when z is real, and therefore coinciding throughout Λ. Furthermore, since for values of z in Λ each term of $R(z)$ when multiplied by z approaches the limit zero when $|z| = \infty$, it follows that $\lim_{|z|=\infty} zR(z) = 0$ from which we conclude that for the same values of z we may write $\lim_{|z|=\infty} \epsilon(z) = 0$.

On the other hand, when z is real and negative we obtain in like manner

$$F(z) = -\frac{\pi}{4z} + \frac{\epsilon(z)}{z}\;; \qquad \lim_{z=-\infty} \epsilon(z) = \lim_{z=+\infty} \epsilon(z) = 0$$

and by reasoning as before we find that this equation holds true for all values of z ($z = 0$ excluded) within any sector lying in the *left* half of the z plane, it being understood that $\lim_{z=-\infty}$ is then replaced by $\lim_{|z|=\infty}$.

Thus, in summary we reach the following conclusion:

According as $-\frac{\pi}{2} + \eta \leqq \arg z \leqq \frac{\pi}{2} - \eta$ *or* $\frac{\pi}{2} + \eta \leqq \arg z \leqq \frac{3\pi}{2} - \eta$, η

† Cf. Osgood, *Encyklopädie*, II 2, p. 21.

being arbitrarily small and positive, we may write when $|z| > 0$

$$F(z) = \begin{cases} \dfrac{\pi}{4z}\left[1 + \epsilon(z)\right] \\ -\dfrac{\pi}{4z}\left[1 - \epsilon(z)\right] \end{cases}; \qquad \lim_{|z| = \infty} \epsilon(z) = 0.$$

For the two sectors above mentioned the function $F(z)$ thus possesses asymptotic developments for which, in the language of § 1, we may take respectively $f_1(z) = 0$, $\phi_1(z) = 1$, $a_{0,1} = 0$, $a_{1,1} = \frac{\pi}{4}$ and $f_2(z) = 0$, $\phi_2(z) = 1$, $a_{0,2} = 0$, $a_{1,2} = -\frac{\pi}{4}$. In case more of the coefficients $a_{n,1}$ or $a_{n,2}$ are desired they may be obtained by using larger values of k in applying (3).

4. *Example* 2. The preceding result may readily be generalized as follows: Let

$$(7) \qquad F(z) = \sum_{n=0}^{n=\infty} \frac{\mu_n}{\lambda_n^2 + z^2}$$

where λ_n, μ_n when considered throughout the continuous domain $n \geqq 0$ are such that (a) *both are continuous*; (b) *both possess continuous first and second derivatives which remain less in absolute value than a constant*;

(c) $\lim\limits_{n=\infty} \lambda_n = +\infty$, $\left|\dfrac{\lambda_n}{n^p}\right| <$ *a constant*, p *being a constant* $> \frac{1}{2}$; (d) $|\mu_n| <$ *a constant.*

The immediate application of (4) then shows that when z is real but different from zero we may write

$$F(z) = \int_0^\infty \frac{\mu_x dx}{\lambda_x^2 + z^2} + \frac{1}{2} \frac{\mu_0}{(\lambda_0^2 + z^2)} + R(z)$$

where $\lim\limits_{z=\pm\infty} zR(z) = 0$. If we now add the fifth hypothesis that (e) *for all values of n sufficiently large ($n \geqq n_0 = a$ constant) the first derivative λ_n' of λ_n shall be positive and always greater than a constant different*

from zero, it appears directly that the integral

$$\int_0^{\infty} \frac{\mu_x dx}{\lambda_x^2 + z^2}$$

will vanish like $\frac{1}{z}$ when $z = \pm \infty$. In fact we shall then have

$$\int_0^{\infty} \frac{\mu_x dx}{\lambda_x^2 + z^2} = \int_0^{n_0} \frac{\mu_x dx}{\lambda_x^2 + z^2} + \int_{n_0}^{\infty} \frac{\mu_x dx}{\lambda_x^2 + z^2}$$

of which the first term in the second member vanishes like $\frac{1}{z^2}$ when $z = \pm \infty$, while the second may be written in the form

$$\theta L \int_{n_0}^{\infty} \frac{\lambda_x' dx}{\lambda_x^2 + z^2} = \frac{\theta L}{z} \Big[\text{arc tan } \frac{\lambda_x}{z} \Big]_{x=n_0}^{x=\infty}$$

where L is the greatest value taken by $\frac{|\mu_n|}{\lambda_n'}$ when $n \geqq n_0$ and where $-1 < \theta < 1$.

In particular, when z is positive we may therefore write

$$F(z) = \frac{a}{z}\Big[1 + \epsilon(z)\Big]; \quad \lim_{z=+\infty} \epsilon(z) = 0$$

where

$$a = \lim_{z=+\infty} z \int_0^{\infty} \frac{\mu_x dx}{\lambda_x^2 + z^2}.$$

We may now carry out the generalizations of this result as in the preceding example, thus reaching the following conclusion:

According as $-\frac{\pi}{2} + \eta \leqq \arg z \leqq \frac{\pi}{2} - \eta$ *or* $\frac{\pi}{2} + \eta \leqq \arg z \leqq \frac{3\pi}{2} - \eta$, η *being arbitrarily small and positive, the function* $F(z)$ *defined by* (7) *may be put into the form*

$$F(z) = \begin{cases} \frac{a}{z}\Big[1 + \epsilon(z)\Big] \\ -\frac{a}{z}\Big[1 - \epsilon(z)\Big] \end{cases}; \quad a = \lim_{z=+\infty} z \int_0^{\infty} \frac{\mu_x dx}{\lambda_x^2 + z^2}; \quad \lim_{|z|=\infty} \epsilon(z) = 0$$

it being assumed throughout that $|z| > 0$ *and that the above indicated conditions* (a), (b), (c), (d) *and* (e) *are satisfied.*

This result is evidently of especial value in all cases where the integral

$$\int_0^\infty \frac{\mu_x dx}{\lambda_x^2 + z^2}$$

is either known or is capable of easy calculation, as in example 1.

It may be added that in order to be assured by the method that the function $\epsilon(z)$ is developable in the precise form called for by (1), where n is supposed to be *arbitrary*, it would be necessary to extend condition (b) to derivatives of *all* orders.

5. *Example* 3. *To obtain asymptotic developments for the function*

$$F(z) = \prod_{n=1}^{n=\infty}\left[1 + \frac{z^2}{n^2}\right].^*$$

As in example 1, this function may be evaluated beforehand and is equal to

$$\frac{e^{\pi z} - e^{-\pi z}}{2\pi z},$$

thus furnishing a check on our subsequent results.

We begin by writing

$$\log F(z) = \sum_{n=1}^{n=\infty} \log\left[1 + \frac{z^2}{n^2}\right] \tag{8}$$

$$= \lim_{m=\infty}\left[\sum_{n=1}^{n=m-1}\log(n^2+z^2) - 2\sum_{n=1}^{n=m-1}\log n\right].$$

From the familiar asymptotic expansion for $\log\{(m-1!)\}$† we have at once

$$-2\sum_{n=1}^{n=m-1}\log n = -2\log\{(m-1)!\} \tag{9}$$

$$= -\log 2\pi - 2(m-\tfrac{1}{2})\log m + 2m + \omega_1(m)\,;\ \lim_{m=\infty}\omega_1(m) = 0\,.$$

We proceed to apply formula (4) to the first summation in the last member of (8), taking for this purpose $f(x) = \log(x^2 + z^2)$ and supposing for the

* Cf. Barnes l. c. (first memoir) § 50.

† Cf. Bromwich, *Infinite Series* (1908), §179.

present that z is real but different from zero. The formula may be applied since the series (2) becomes

$$\Omega_1(m) = \sum_{n=1}^{n=\infty} \int_0^1 \left\{ \frac{d^2}{dx^2} \log(x^2+z^2) \right\}_{x=m+n-t} \phi_2(t)dt$$

and this is readily seen to be convergent when we apply the first law of the mean for integrals to its nth term (cf. example 1). Likewise we see that $\lim_{m=\infty} \Omega_1(m) = 0$.

Formula (4) thus gives

$$(10) \qquad \sum_{n=1}^{n=m-1} \log(n^2+z^2) = \int_1^m \log(m^2+z^2)dm - \tfrac{1}{2}\log(m^2+z^2) + \Omega_1(m) + \\ + \tfrac{1}{2}\log(1+z^2) + R(z)$$

where

$$(11) \qquad R(z) = -\Omega_1(1) = -\sum_{n=1}^{n=\infty} \int_0^1 \left\{ \frac{d^2}{dx^2} \log(x^2+z^2) \right\}_{x=n-t+1} \phi_2(t)dt.$$

By combining relations (8), (9) and (10) and then making use of the relation

$$\int_1^m \log(m^2+z^2)dm = m\log(m^2+z^2) - 2m + 2z \text{ arc tan } \frac{m}{z} \\ - \log(1+z^2) + 2 - 2z \text{ arc tan } \frac{1}{z},$$

we obtain

$$\log F(z) = -\log 2\pi - \tfrac{1}{2}\log(1+z^2) - 2z \text{ arc tan } \frac{1}{z} + 2 + R(z) + \\ \lim_{m=\infty} \left[(m-\tfrac{1}{2})\log\left(1+\frac{z^2}{m^2}\right) + 2z \text{ arc tan } \frac{m}{z} + \omega_1(m) + \Omega_1(m) \right].$$

Moreover, we have

$$\lim_{m=\infty} (m-\tfrac{1}{2})\log\left(1+\frac{z^2}{m^2}\right) = 0, \qquad \lim_{m=\infty} \omega_1(m) = 0, \qquad \lim_{m=\infty} \Omega_1(m) = 0$$

and, supposing at first that z is *positive*, we shall have

$$\lim_{m=\infty} 2z \text{ arc tan } \frac{m}{z} = \pi z$$

and hence

$$(12)\quad \log F(z) = -\log 2\pi z + \pi z - \tfrac{1}{2}\log\left(1+\frac{1}{z^2}\right) + 2\left(1 - z \operatorname{arc\,tan}\frac{1}{z}\right) + R(z).$$

This result may now be generalized, as in example 1, for all values of z ($z = 0$ excluded) for which $-\frac{\pi}{2} + \eta \gtreqless \arg z \gtreqless \frac{\pi}{2} + \eta$, and for such values it appears directly from (11) that $\lim_{|z|=\infty} R(z) = 0$. Also, for the same values of z we have

$$\lim_{|z|=\infty} \log\left(1+\frac{1}{z^2}\right) = 0, \qquad \lim_{|z|=\infty}\left(1 - z \operatorname{arc\,tan}\frac{1}{z}\right) = 0,$$

and hence

$$\log F(z) = -\log 2\pi z + \pi z + \eta(z)\,; \qquad \lim_{|z|=\infty} \eta(z) = 0.$$

On the other hand, when z is negative we have, instead of (12),

$$\log F(z) = -\log(-2\pi z) - \pi z - \tfrac{1}{2}\log\left(1+\frac{1}{z^2}\right) + 2\left(1 - z \operatorname{arc\,tan}\frac{1}{z}\right) + R(z).$$

Thus we reach the following result:

According as $-\frac{\pi}{2} + \eta \gtreqless \arg z \gtreqless \frac{\pi}{2} - \eta$ *or* $\frac{\pi}{2} + \eta \gtreqless \arg z \gtreqless \frac{3\pi}{2} - \eta$, *$\eta$ being arbitrarily small and positive, we may write when* $|z| > 0$

$$F(z) = \begin{cases} \dfrac{e^{\pi z}}{2\pi z}\Big[1 + \epsilon(z)\Big] \\[2ex] -\dfrac{e^{-\pi z}}{2\pi z}\Big[1 - \epsilon(z)\Big] \end{cases} \qquad \lim_{|z|=\infty} \epsilon(z) = 0.$$

6. *Example* 4. In generalization of the last result we shall merely point out certain consequences of the method as applied to the function

$$F(z) = \prod_{n=1}^{n=\infty}\left[1 + \frac{z^2}{\lambda_n^2}\right]$$

where λ_n when considered throughout the continuous domain $n \geqq 1$ satisfies the following conditions: (*a*) *never vanishes*; (*b*) *is continuous*; (*c*) *has*

continuous first and second derivatives which remain less in absolute value than a constant; and (*d*) $\lim\limits_{n=\infty} \dfrac{\lambda_n}{n} =$ *a constant* $\neq 0$.

We begin by writing

$$(13) \qquad \log F(z) = \lim_{m=\infty} \left[\sum_{n=1}^{n=m-1} \log(\lambda_n^2 + z^2) - 2 \sum_{n=1}^{n=m-1} \log \lambda_n \right].$$

From (4) we have

$$(14) \quad -2 \sum_{n=1}^{n=m-1} \log \lambda_n = c - \int_1^m \log \lambda_m^2 dm + \log \lambda_m + \omega_1(m)\,; \quad \lim_{m=\infty} \omega_1(m) = 0$$

where

$$(15) \qquad c = -\log \lambda_1 + 2 \sum_{n=1}^{n=\infty} \int_0^1 \left\{ \frac{d^2}{dx^2} \log \lambda_x \right\}_{x=n-t+1} \phi_2(t) dt.$$

Again, if z be real but different from zero, the application of (4) to the first summation on the right in (13) gives

$$(16) \qquad \sum_{n=1}^{n=m-1} \log(\lambda_n^2 + z^2) = \int_1^m \log(\lambda_m^2 + z^2) dm - \tfrac{1}{2} \log(\lambda_m^2 + z^2)$$

$$+ \Omega_1(m) + \tfrac{1}{2} \log(\lambda_1^2 + z^2) + R(z)\,; \qquad \lim_{m=\infty} \Omega_1(m) = 0$$

where

$$(17) \qquad R(z) = -\sum_{n=1}^{n=\infty} \int_0^1 \left\{ \frac{d^2}{dx^2} \log(\lambda_x^2 + z^2) \right\}_{x=n-t+1} \phi_2(t) dt.$$

Moreover, by an integration by parts we obtain

$$(18) \qquad \int_1^m \log(\lambda_m^2 + z^2) dm = m \log(\lambda_m^2 + z^2) - \log(\lambda_1^2 + z^2)$$

$$- 2 \int_1^m \frac{m \lambda_m \lambda_m' dm}{\lambda_m^2 + z^2}$$

where λ_m' denotes the derivative of λ_m with respect to m.

Let us now make use of (14) and (15) in relation (13) and subsequently expand by means of (18) the second term in the second member of (14) and the first term in the second member of (16).

Upon placing $m = \infty$ in the result we obtain

$$\log F(z) = K - \tfrac{1}{2}\log(\lambda_1^2 + z^2) + 2z^2 \int_1^\infty \frac{m\lambda_m' dm}{\lambda_m(\lambda_m^2 + z^2)} + R(z)$$

where $K = c + 2 \log \lambda_1$.

If we now introduce the additional hypothesis that *for all values of n greater than some constant n_0 we shall have $\lambda_n' \geqq 0$*, we find directly in the manner already indicated in example 2 that according as z is positive or negative we shall have

$$\log F(z) = \begin{cases} K - \log z + 2a_2 z + \epsilon(z) \\ K - \log(-z) - 2a_2 z - \epsilon(z) \end{cases}; \quad \lim_{z=+\infty} \epsilon(z) = \lim_{z=-\infty} \epsilon(z) = 0$$

where

$$a_2 = 2 \lim_{z=+\infty} z\int_1^\infty \frac{m\lambda_m' dm}{\lambda_m(\lambda_m^2 + z^2)} \tag{19}$$

and, if we now proceed as in the previous examples, we may generalize this result into the following:

According as $-\dfrac{\pi}{2} + \eta \leqq arg\ z \leqq \dfrac{\pi}{2} - \eta$ *or* $\dfrac{\pi}{2} + \eta \leqq arg\ z \leqq \dfrac{3\pi}{2} - \eta$,

η being arbitrarily small and positive, we may write when $|z| > 0$

$$F(z) = \begin{cases} \dfrac{a_1 e^{a_2 z}}{z}\Big[1 + \epsilon(z)\Big] \\ \dfrac{-a_1 e^{-a_2 z}}{z}\Big[1 - \epsilon(z)\Big] \end{cases}; \qquad \lim_{|z| = \infty} \epsilon(z) = 0$$

where $a_1 = e^K$ and a_2 are constants.

As a noteworthy special consequence of this result, obtained from it by making the substitution $z = iz$ ($i = \sqrt{-1}$) and expanding the resulting exponential functions by De Moivre's theorem, we note the following:

Let $\Phi(z)$ *be any function of the complex variable z expressible in the form*

$$\Phi(z) = \prod_{n=1}^{n=\infty} \left[1 - \frac{z^2}{\lambda_n^2}\right]$$

in which λ_n when considered throughout the continuous domain $n \geqq 1$ satisfies the following conditions: (a) never vanishes; (b) is continuous; (c) has

continuous first and second derivatives of which the first eventually becomes and remains positive, while both always remain less in absolute value than a constant; (*d*) $\lim_{n=\infty} \frac{\lambda_n}{n}$ *= a constant different from zero.*

According as $\eta \gtreqless \arg z \gtreqless \pi - \eta$ *or* $-\pi + \eta \gtreqless \arg z \gtreqless -\eta$, η *being arbitrarily small and positive, we may write when* $|z| > 0$

$$\Phi(z) = \begin{cases} \frac{a_1}{z}[\sin a_2 z - i \cos a_2 z][1 + \epsilon(z)] \\ \frac{a_1}{z}[\sin a_2 z + i \cos a_2 z][1 - \epsilon(z)] \end{cases}; \quad \lim_{|z|=\infty} \epsilon(z) = 0$$

where a_1 *and* a_2 *are constants.*

7. It will be readily perceived that the method employed in the preceding examples has a wide field of applicability. While we shall not attempt to carry the subject further, it may be noted that information may, in general, be thus obtained concerning the behavior for large values of $|z|$ of any function $F(z)$ defined by a series of the form

$$F(z) = \sum_{n=a}^{n=\infty} f_n(z)$$

when, by taking k sufficiently large, it can be shown that the series

$$\Omega_k(m) = \sum_{n=1}^{n=\infty} \int_0^1 \left\{ \frac{d^{2k}}{dx^{2k}} f_x(z) \right\}_{x=m+n-t} \phi_{2k}(t)dt; \qquad m \geqq a$$

is convergent for all values of z in a certain real region, while for the same values of z it can be shown also that $\lim_{m=\infty} \Omega_k(m) = 0$. In such cases we may apply at once formula (3) after which, upon letting $m = \infty$, we obtain an expression for $F(z)$ which yields the desired asymptotic behavior when z is confined to the preassigned real region. By studying the properties of the same expression, we may usually generalize the result so as to obtain the asymptotic behavior for one or more sectors of the z complex plane.

UNIVERSITY OF MICHIGAN
ANN ARBOR, MICH.

THE INTEGRAL ROOTS OF CERTAIN INEQUALITIES

By W. H. Jackson

1. Introductory. Let the residue, mod 1, of any number x be denoted by $F(x)$. Let d be any number and β, β' be positive numbers less than 1.

The following paper is concerned with integral solutions, Y, of the inequalities

$$\beta > F(Yd) > \beta'. \tag{1}$$

If d is a commensurable number, P/Q, P and Q being positive integers prime to each other, and all positive integral roots $[Y_r]_{r=1}^{r=k}$ less than Q have been found by trial, the complete solution is given by

$$Y = Y_1, Y_2, \cdots Y_k, \text{ mod } Q. \tag{2}$$

That is, the series of roots of (1), arranged in order of magnitude possesses a period Q.

If d is not a commensurable number, this period disappears but is replaced by a quasi-periodicity, the regularity of which is marred by gaps at certain points to be found later.

Further, even when d is commensurable, this quasi-periodic structure may be found within the regular period Q. This is illustrated by the following example:

Let
$$\tfrac{3}{4} > F(Y\tfrac{7}{58}) > \tfrac{1}{4}. \tag{3}$$

The solutions $Y_1, Y_2, \cdots Y_k$ are as follows:

3, 4, 5, 6,	3, 4, 5, 6
11, 12, 13, 14,	8 + 3, 4, 5, 6
19, 20, 21, 22,	2·8 + 3, 4, 5, 6
27, 28, 29, 30, 31	25 + 2, 3, 4, 5, 6
36, 37, 38, 39,	25 + 8 + 3, 4, 5, 6
44, 45, 46, 47,	25 + 2·8 + 3, 4, 5, 6
52, 53, 54, 55.	2·25 + 2, 3, 4, 5.

The second method of arrangement shows clearly that within the regular period of $58 (= 2 \cdot 25 + 8)$ there is a clearly marked quasi-periodicity, like that with which the recurrence of eclipses has made us familiar. It seems

hardly correct to replace the term quasi-periodicity by the shorter word periodicity because neither the period of repetition nor the group repeated is quite permanent.

There is a primary group, 3, 4, 5, 6, to which the term 2 may be added, or from which the term 6 may be omitted, and this we may denote by S_1.

There is a secondary group which we will call S_2, which consists of the terms

$$S_1, \qquad S_1 + 8, \qquad S_1 + 2\cdot 8,$$

where it is to be understood that the final number is to be added to each member of the group S_1.

The final group S, which is, in this case of the third order, consists of the terms

$$S_2, \qquad S_2 + 25, \qquad S_1 + 2\cdot 25.$$

This group is repeated with perfect regularity.

It is the object of the present paper to develop a method by which this periodic structure can be studied in detail. Actually to exhibit this structure, built up of one period within another, by a system of formulas which includes all possible cases is too complicated a result to be reproduced here.

It is of interest to note that the results are not limited to *commensurable* values of d, β or β'.

The solution of the present problem was attempted in order to answer questions raised by the paper on shadow rails which immediately follows it.

The writer has found no references pertinent to the subject.

2. Notation. As may be readily guessed, it is the expansion of d as a simple continued fraction which is the initial step from which all else follows.

Suppose d to be expanded in a simple continued fraction as below.

Let $$d = a_1 + \frac{1}{a_2 +}\,\frac{1}{a_3 +}\cdots\frac{1}{a_n + e_n}, \tag{4}$$

where the a's are positive integers, $0 < e_n < 1$,

and $$e_{n-1} = \frac{1}{a_n + e_n}. \tag{5}$$

The coefficients a_r are most readily determined by applying the ordinary process for finding the highest common factor of two numbers to the numbers d and 1. In the case where d is a commensurable number, P/Q, it is more

convenient to apply the process to P and Q. The equations thus obtained will be the same as those which follow, multiplied throughout by Q.

Thus let

$$
\begin{aligned}
d &= a_1 \cdot 1 + w_1, & 0 &< w_1 < 1,\\
1 = w_0 &= a_2 \cdot w_1 + w_2, & 0 &< w_2 < w_1,\\
w_1 &= a_3 \cdot w_2 + w_3, & 0 &< w_3 < w_2,\\
&\cdots\cdots\\
w_{n-2} &= a_n \; w_{n-1} + w_n, & 0 &< w_n < w_{n-1}.
\end{aligned} \qquad (6)
$$

The quantities w_r, not considered in the usual treatment of continued fractions, are fundamental in the present discussion.*

Let the nth convergent be denoted by p_n/q_n. Since this is obtained from the $(n-1)$th convergent by the substitution of $a_{n-1} + 1/a_n$ for a_{n-1}, it is easily seen that the following equations are true:

$$
\begin{aligned}
p_1 &= a_1 \cdot 1 + 0, & q_1 &= a_1 \cdot 0 + 1,\\
p_2 &= a_2 \cdot p_1 + p, & q_2 &= a_2 \cdot q_1 + 0,\\
p_3 &= a_3 \cdot p_2 + p_1, & q_3 &= a_3 \cdot q_2 + q_1,\\
&\cdots\cdots & &\cdots\cdots\\
p_n &= a_n \cdot p_{n-1} + p_{n-2}, & q_n &= a_n \cdot q_{n-1} + q_{n-2}.
\end{aligned} \qquad (7)
$$

Further, equations (6) may be put into the same form as those just written:

$$
\begin{aligned}
-w_1 &= a_1 \cdot 1 - d,\\
w_2 &= a_2 \cdot (-w_1) + w_0,\\
-w_3 &= a_3 \cdot w_2 - w_1,\\
&\cdots\cdots\\
(-1)^n w_n &= a_n \cdot (-1)^{n-1} w_{n-1} + (-1)^{n-2} w_{n-2}.
\end{aligned} \qquad (8)
$$

Lastly, if in equations (7) we multiply each of the equations involving the convergents q_r by $-d$ and add to the corresponding equations involving p_r, we find by comparison with equations (8) that

$$(-1)^n w_n = p_n - dq_n,$$

or

$$q_n d = p_n + (-1)^{n+1} w_n. \qquad (9)$$

---

* It follows by comparison with equations (4), (5) that

$$e_1 = w_1, \quad e_2 = w_2/w_1, \ldots e_n = w_n/w_{n-1}, \quad \text{whence} \quad w_n = e_1 e_2 e_3 \ldots e_n.$$

The mode of construction of equations (6) ensures that the quantities w_r form a sequence of positive quantities approaching zero as a limit. Equation (9) shows that they are fundamental in calculating the residues, mod 1, of integral multiples of d.

It will be shown (Theorem B) that any positive quantity β, less than 1, can be expressed as the sum of either a finite series or a convergent infinite series of integral multiples of the quantities w_r. If in this series $(-1)^{n+1} w_n$ is replaced by q_n, we obtain a corresponding series, divergent if infinite, such that if B_n denote the sum of its first n terms,

$$\lim_{n=\infty} F(B_n d) = \beta.$$

The converse process of finding $F(Bd)$, when B is given, will in general best be accomplished directly by multiplication rather than indirectly by first expanding B in a series of multiples of the convergents q_r.

3. The Fundamental Theorems. Let $b_1, b_2, \cdots b_n$ be positive integers and let

$$B_n = \sum_{r=1}^{n} b_r q_r, \tag{10}$$

or, what comes to the same thing, let

$$B_n = b_n q_n + B_{n-1}. \tag{11}$$

THEOREM A. *Any positive integer can be uniquely expressed as a series* (10) *by means of the inequalities*

$$q_{n+1} > B_n \geqq 0. \tag{12}$$

Further, the coefficients, b_n *so determined satisfy the inequalities*

$$a_{n+2} \geqq b_{n+1} \geqq 0, \quad \textit{when} \quad b_n = 0, \qquad n \neq 0,$$

and $$a_{n+2} - 1 \geqq b_{n+1} \geqq 0, \quad \textit{when} \quad n = 0 \quad \textit{or} \quad b_n \neq 0, \quad n \neq 0. \tag{14}$$

If inequalities (12) hold good for B_n, it follows from equations (7) that two cases arise in which either

$$\text{(i)} \qquad q_{n+1} > B_n \geqq a_{n+1} q_n, \tag{15}$$

or

$$\text{(ii)} \qquad a_{n+1} q_n > B_n \geqq 0. \tag{16}$$

In each case assume that inequalities (12) hold good for B_{n-1}. It then follows from (11) that

$$q_n > B_n - b_n q_n \geqq 0. \tag{17}$$

That is, b_n is the greatest integer contained in B_n/q_n, and the coefficients b_n are uniquely determined.

In case (i), therefore, inequalities (15) and (17) give

$$q_{n+1} + q_n > (b_n + 1)q_n > a_{n+1}q_n.$$

Hence, from (7)

$$a_{n+1}q_n + q_{n-1} > b_nq_n > (a_{n+1} - 1)q_n,$$

that is

$$b_n = a_{n+1}. \tag{18}$$

Further, if inequalities (12) hold for B_{n-2}, we see from (17) that

$$q_{n-1} > B_{n-1} - b_{n-1}q_{n-1} \geqq 0,$$

and from a second application of (11),

$$B_n - a_{n+1}q_n - b_{n-1}q_{n-1} \geqq 0.$$

From (7) and (15)

$$(1 - b_{n-1})q_{n-1} > 0,$$

whence

$$b_{n-1} = 0. \tag{19}$$

In case (ii), it follows similarly from (16) and (17) that

$$a_{n+1}\, q_n > b_nq_n > -\, q_n,$$

whence

$$a_{n+1} - 1 \geqq b_n \geqq 0. \tag{20}$$

Equations (18), (19), (20) show that the coefficients b_n satisfy inequalities (14) when $n \neq 0$. It follows directly from (7) and (12) that, also when $n = 0$, inequalities (14) are satisfied.

Corollary. Conversely, if inequalities (14) are satisfied it follows that inequalities (12) are satisfied and hence that there is only one expansion of B_n in which the coefficients satisfy (14).

The proof is as follows. Assume that

$$q_{n+1} > B_n \geqq 0.$$

If

$$(a_{n+2} - 1) \geqq b_{n+1} \geqq 0,$$

$$(a_{n+2} - 1)\, q_{n+1} \geqq b_{n+1}q_{n+1} \geqq 0,$$

and therefore

$$q_{n+2} - q_{n+1} \geqq B_{n+1} \geqq 0, \qquad \text{from (7) and (11).}$$

If

$$b_{n+1} = 0, \quad b_{n+2} = a_{n+3},$$

from (11) $$B_{n+2} = a_{n+3}q_{n+2} + B_n,$$

and therefore $$a_{n+3}q_{n+2} + q_{n+1} > B_{n+2} > 0,$$

that is $$q_{n+3} > B_{n+2} > 0.$$

That is, if inequalities (12) are true for n, they are true for $n+1$. But from (7) and (14) it follows that they are true when $n = 1$. They are therefore satisfied for all positive integral values of n, which proves the theorem.

It follows from equations (9) and (10) that

$$B_n d = \sum_{r=1}^{n} b_r p_r + \sum_{r=1}^{n} (-1)^{r+1} b_r w_r. \tag{21}$$

That is, the residue, mod 1, of $B_n d$ can be uniquely expressed as a series of multiples of the quantities w_r, and the coefficients b_r, satisfy (14). This raises the converse question, can any positive quantity β, less than unity be expressed as the sum of such a series, and so determine a multiplier B such that

$$F(Bd) = \beta.$$

First it is necessary to consider the nature of the convergence of such a series.

Let $$\beta_n = \sum_{r=1}^{n} (-1)^{r+1} b_r w_r. \tag{22}$$

Let the first odd and even coefficients which are not zero be b_{2u-1} and b_{2s} respectively.

Equations (6) enable us to write

$$w_n > a_{n+2} w_{n+1}. \tag{23}$$

It is convenient to note the following consequences of inequalities (14) and (23), when use is made of equations (6).

(i) If $s \geqq u = 1$, $$1 - w_1 > \beta_n > w_2,$$

and if $s \geqq u \geqq 2$, $$w_{2u-2} > \beta_n > w_{2u}. \tag{24}$$

(ii) If $u \geqq s + 1$, $$-w_{2s+1} > \beta_n > -w_{2s-1}. \tag{25}$$

(iii) If $n > r$, $b_{r+1} \neq a_{r+2}$,

$$w_r - w_{r+1} > (-1)^r(\beta_n - \beta_r) > -w_{r+1}. \tag{26}$$

(iv) If $n > r$, $b_{r+1} = a_{r+2}$,

$$w_r > (-1)^r(\beta_n - \beta_r) > w_r - w_{r+1}. \tag{27}$$

(v) In all cases, $1 - w_1 > \beta_n > - w_1$.

The above inequalities are written out for the case in which d is incommensurable. When d is commensurable a sign of equality must be inserted either at the upper or lower limit. The limit at which it must be placed depends on whether the last a_n which is not zero has an odd or even suffix.

It is clear that the series β_n is convergent for all possible values of the coefficients b_r, lying within the limits prescribed by (14).

Of the above results, inequalities (26), on account of their generality, form the best starting point for the expansion of any positive number less than 1 in a series β_n.

Let
$$\beta - \beta_0 = \beta_n + (-1)^n \rho_n, \tag{28}$$
where $\beta_0 = 0$ or 1 according as $\beta <$ or $\gtreqless 1 - w_1$. (29)

It follows from (22) and (28) that

$$\rho_n + \rho_{n+1} = b_{n+1} w_{n+1}. \tag{30}$$

Inequalities (26) suggest the following as a means of determining the expansion (28):

$$w_n > \rho_n + w_{n+1} \geqq 0. \tag{31}$$

THEOREM B. *Any positive quantity β, less than 1, can be expanded uniquely in the series (28), if each positive integer b_n is determined for successive values of n, when possible, from the inequality for ρ_n in (31) and is otherwise zero. Further, the values of b_n so obtained satisfy (14).*

Making use of equations (30), we may write (31) in the form

$$w_n > b_n w_n - \rho_{n-1} + w_{n+1} \geqq 0. \tag{32}$$

These inequalities determine b_n as the least integer not less than

$$(\rho_{n-1} - w_{n+1}) / w_n.$$

Again, from (31), we assume

$$w_{n-1} > \rho_{n-1} + w_n \geqq 0,$$

whence, from (6),

$$(a_{n+1} - 1) w_n > \rho_{n-1} - w_{n+1} \geqq - w_n - w_{n+1},$$

and, therefore

$$a_{n+1} - 1 \geqq b_n \geqq -1. \tag{33}$$

If

$$w_{n+1} \geqq \rho_{n-1} + w_n \geqq 0, \tag{34}$$

b_n as determined by (32) would be negative, -1 in fact; we write $b_n = 0$ and in this case we find by a double application of equations (30) to inequalities (31) for ρ_{n+1} that

$$w_{n+1} > b_{n+1} w_{n+1} + \rho_{n-1} + w_{n+2} \geqq 0,$$

which determines b_{n+1} as the least integer not less than $(-\rho_{n-1} - w_{n+2})/w_{n+1}$.

Further, from (34),

$$2w_{n+1} > (b_{n+1} + 1) w_{n+1} - w_n + w_{n+2} \geqq 0.$$

Comparison with equations (6) shows that

$$2w_{n+1} > (b_{n+1} + 1 - a_{n+2}) w_{n+1} \geqq 0,$$

that is

$$b_n = 0, \quad b_{n+1} = a_{n+2} \quad \text{or} \quad a_{n+2} - 1. \tag{35}$$

Hence if one value of ρ_n is found satisfying inequalities (31), the series is uniquely determined for any number of terms greater than n.

But when $\beta < 1 - w_1$, $\rho_0 = \beta$, and we have

$$w_0 > \rho_0 + w_1 > 0.$$

And when $\beta \geqq 1 - w_1$, $\rho_0 = \beta - 1$, and again

$$w_0 > \rho_0 + w_1 \geqq 0.$$

That is, the series is in all cases uniquely determined.

Lastly, it is clear from inequalities (33), (35) that the coefficients b_n, satisfy inequalities (14).

Corollary. If ρ_n does not satisfy inequalities (31), it follows from (34) by the means of (30) that, since $b_n = 0$,

$$w_n + w_{n+1} \geqq \rho_n + w_{n+1} \geqq w_n. \tag{36}$$

4. Applications. As a half-way step towards finding the integral roots of inequalities (1), it is convenient to formulate the conditions for

$$F(Yd) \gtrless \beta,$$

where β has the value assigned in (28).

Let Y be expanded by Theorem A into the form

$$Y = Y_n + S_n, \tag{37}$$

where $$Y_n = \sum_{r=l}^{n} y_r q_r, \quad S_n = \sum_{r=n}^{n+m} y_r q_r. \tag{38}$$

It follows from equations (9), that if

$$\eta_n = \sum_{r=l}^{n} (-1)^{r+1} y_r w_r, \quad (-1)^n \sigma_n = \sum_{r=n}^{n+m} (-1)^{r+1} y_r w_r, \tag{39}$$

then $$F(Yd) = \eta_0 + \eta_n + (-1)^n \sigma_n, \tag{40}$$

where η_0 is 0 or 1 as l is odd or even, as may be seen by referring to inequalities (24), (25).

It follows from equations (6) that starting out from the critical value $1 - w_1$, the interval from 0 to 1 is completely made up of the intervals separating

$$1 - w_1,\ w_2,\ w_4,\ w_6 \cdots \text{ on the one hand, and}$$

$$1 - w_1,\ 1 - w_3,\ 1 - w_5, \cdots \text{ on the other.}$$

Two cases are best considered separately, in which either

$$\text{I,} \quad \beta < 1 - w_1, \qquad \text{or} \qquad \text{II,} \quad \beta \geqq 1 - w_1.$$

I. In the first case, let β be contained in the interval $w_{2u-2} w_{2u}$. From inequalities (24) (25)

$$F(Yd) > \beta, \quad \text{if} \quad l = 2k - 1,\ k < u, \quad \text{or if} \quad l = 2s,$$

$$F(Yd) < \beta, \quad \text{if} \quad l = 2k - 1,\ k > u. \tag{41}$$

If $l = 2u - 1$, write $\beta = b_{2u-1} w_{2u-1} - \rho_{2u-1}$, where, from (31),

$$w_{2u-1} > \rho_{2u-1} + w_{2u} \geqq 0. \tag{42}$$

$$F(Yd) = y_{2u-1} w_{2u-1} - \sigma_{2u-1},$$

where, from (26) and (27),

$$w_{2u-1} > \sigma_{2u-1} + w_{2u} > 0. \tag{43}$$

It follows from (42) and (43) that

$$w_{2u-1} > |\rho_{2u-1} - \sigma_{2u-1}| \geqq 0,$$

whence, if $l = 2u - 1$, $F(Yd) \gtrless \beta$ according as $y_{2u-1} \gtrless b_{2u-1}$. (44)

$$\text{If } \left[y_r = b_r\right]_{r=2u-1}^{r},\ y_{r+1} \neq b_{r+1},$$

$$\beta = \beta_r + (-1)^r (b_{r+1} w_{r+1} - \rho_{r+1}),$$

$$F(Yd) = \beta_r + (-1)^r (y_{r+1} w_{r+1} - \sigma_{r+1}).$$

Three cases now arise, in which

(i) $$w_{r+1} > \rho_{r+1} + w_{r+2} \geqq 0,$$

$$w_{r+1} > \sigma_{r+1} + w_{r+2} > 0,$$

whence $$w_{r+1} > |\rho_{r+1} - \sigma_{r+1}| > 0.$$

(ii) $$b_{r+1} = 0,\ w_{r+1} + w_{r+2} \geqq \rho_{r+1} + w_{r+2} \geqq w_{r+1},$$

$$y_{r+1} > b_{r+1},\quad w_{r+1} > \sigma_{r+1} + w_{r+2} > 0,$$

whence $$\rho_{r+1} - \sigma_{r+1} > 0.$$

(iii) $$b_{r+1} > y_{r+1},\ w_{r+1} > \rho_{r+1} + w_{r+2} \geqq 0,$$

$$y_{r+1} = 0,\ w_{r+1} + w_{r+2} > \sigma_{r+1} + w_{r+2} > w_{r+1},$$

whence $$\sigma_{r+1} - \rho_{r+1} > 0.$$

Hence in all cases, when $[y_r = b_r]^r_{r=2u-1}$, $y_{r+1} \neq b_{r+1}$,

$$F(Yd) \gtrless \beta \quad \text{as} \quad (-1)^r y_{r+1} \gtrless (-1)^r b_{r+1}. \tag{45}$$

II. In the second case, let β be contained in the interval $1 - w_{2s-1}$, $1 - w_{2s+1}$. Results corresponding exactly to those just proved hold in this case also.

$$F(Yd) < \beta,\ \text{if } l = 2k,\ k < s \ \text{ or if } l = 2u - 1,$$

$$F(Yd) > \beta,\ \text{if } l = 2k,\ k > s. \tag{46}$$

If $$l = 2s,\quad F(Yd) \gtrless \beta \quad \text{according as} \quad y_{2s} \gtrless b_{2s}. \tag{47}$$

If $$[y_r = b_r]^r_{r=2s},\quad y_{r+1} \neq b_{r+1},$$

$$F(Yd) \gtrless \beta \quad \text{according as} \quad (-1)^r y_{r+1} \gtrless (-1)^r b_{r+1}. \tag{48}$$

We are now in a position to solve inequalities (1).

Suppose that $$1 - \beta' = \beta_0' + \beta_n' + (-1)^n \rho_n', \tag{49}$$

where β_0', β_n', ρ_n', b_n', B_n' have values analogous to those of β_0, β_n, ρ_n, b_n, B in equations (29), (28), (22), (10).

Further let $\beta'' = \beta - \beta'$.

It now follows from (40) that if Y be any root of (1),

$$1 - \beta_0' + \beta'' + \rho_{2n+1}' > F(Yd) + F(B_{2n+1}'d) > 1 - \beta_0' + \rho_{2n+1}'.$$

That is
$$\beta'' + \rho'_{2n+1} > F(Y'd) > \rho'_{2n+1}, \tag{50}$$
where
$$Y' = Y + B'_{2n+1}.$$
Consider now the inequalities
$$\beta'' > F(Y'd) > 0. \tag{51}$$

By making n large, as many integral roots as we please of (51) can be made equal to corresponding roots of (50) and can therefore be made to exceed a corresponding root of (1) by the constant B'_{2n+1}.

The solutions of (51) have been enumerated in inequalities (41)to(48). These formulas, therefore, provide the solution of inequalities (1).

Let us now return to the example already considered, in which $d = 7/58$. We obtain on applying the method of finding the highest common factor of 7 and 58, the following equations, corresponding to equations (6) and (7).

$$\begin{array}{lll}
7 = 0 \cdot 58 + 7, & p_1 = 0 \cdot 1 + 0 = 0, & q_1 = 0 \cdot 0 + 1 = 1, \\
58 = 8 \cdot 7 + 2, & p_2 = 8 \cdot 0 + 1 = 1, & q_2 = 8 \cdot 1 + 0 = 8, \\
7 = 3 \cdot 2 + 1, & p_3 = 3 \cdot 1 + 0 = 3, & q_3 = 3 \cdot 8 + 1 = 25, \\
2 = 2 \cdot 1 + 0. & P = 2 \cdot 3 + 1 = 7. & Q = 2 \cdot 25 + 8 = 58.
\end{array}$$

That is
$$\frac{7}{58} = 0 + \frac{1}{8} + \frac{1}{3} + \frac{1}{2},$$
$$58\,w_1 = 7, \quad 58\,w_2 = 2, \quad 58\,w_3 = 1, \quad w_4 = 0.$$
We find also that
$$\frac{3}{4} = \frac{43 \cdot 5}{58} = 6\,w_1 - 0 \cdot w_2 + 2\,w_3 - \frac{\cdot 5}{58} \tag{52}$$
$$\frac{1}{4} = \frac{14 \cdot 5}{58} = 2\,w_1 - 0 \cdot w_2 + \,w_3 - \frac{\cdot 5}{58}. \tag{53}$$

Hence the roots of (3), less than Q, are to be found amongst the numbers
$$\begin{bmatrix} 6 \\ : \\ 2 \end{bmatrix} q_1 + \begin{bmatrix} 2 \\ \cdot \\ 0 \end{bmatrix} q_2 + \begin{bmatrix} 2 \\ \\ 0 \end{bmatrix} q_3,$$
namely,
$$[2, 3, 4, 5, 6] + [0, 1, 2]8 + [0, 1, 2]25,$$
as may be verified by referring to the table of roots following (3). From (53) the number 2 in the first term only appears in conjunction with 0 in the

second and 1, 2 in the third term. From (52), the number 6 in the first term is omitted when it would appear in conjunction with 0 in the second term and 2 in the third. From (14) only 0 in the second term can appear along with 2 in the third. Whence the number of terms is $4 \cdot 3 \cdot 3 + 2 - 1 - 4 \cdot 2 = 29$ which agrees with the table above. But the general law of formation is more easily followed if we construct the inequalities corresponding to (51). In this case

$$\beta'' = \frac{1}{2} = \frac{29}{58} = 4w_1 - 0 \cdot w_2 + w_3.$$

From (52) $$B'_3 = 6 + 2q_3 = 56, \quad Y' = 56 + Y,$$

$$\frac{1}{2} \geqq F\left(Y' \frac{7}{58}\right) > 0.$$

The various groups are now as follows:

$$S_1 = [1,2,3,4], \ S_1' = [0,1,2,3],$$

$$S_2 = S_1 + [0,1,2]8, \ S_2' = S_2, 0,$$

$$S_3 = S_2 + [0]25, \ S_2' + [1]25, \ S_1' + [2]25, \tag{54}$$

whence

$$Y - 2 \equiv Y' \equiv \left\{\begin{array}{rrrrr} & 1, & 2, & 3, & 4 \\ & 9, & 10, & 11, & 12 \\ & 17, & 18, & 19, & 20 \\ 25, & 26, & 27, & 28, & 29 \\ & 34, & 35, & 36, & 37 \\ & 42, & 43, & 44, & 45 \\ 50, & 51, & 52, & 53, & \end{array}\right\} \text{mod } 58,$$

which is in agreement with previous results.

The above equations (54) connecting S_1, S_2, S_3 have been written out above, because, although it would take up too much space to exhibit the laws of formation of these groups in general, still those already given are quite typical of the results obtained in the general case.

One other result which is a special case of another general theorem may be noted. The group S_1 is repeated 7 times in S_3, and therefore its average period of repetition is 58/7, which is the same thing as $1/w_1$. It is also the same thing as $1/d$, but $1/w_1$ is chosen, because the theorem states that in gen-

eral, this average period is $1/w_{2n-1}$, where b''_{2n-1} is the first coefficient in the expansion of β'' which is not zero.

In conclusion, we shall prove the theorem which determines the smallest interval between two values of Y. Let us denote this smallest interval by L.

Let $$w_{2n-2} > \beta'' > w_{2n}. \tag{55}$$

If $\beta'' > 1 - w_1$, $L = q_1$ by (46). Otherwise, by (24), (31),

$$\beta'' = b''_{2n-1} w_{2n-1} - \rho_{2n-1}. \tag{56}$$

Also, by (41), (44) and (45)

$$F(Y'd) = y_{2n-1} w_{2n-1} - \sigma_{2n-1}, \qquad b''_{2n-1} \geqq y_{2n-1} \geqq 0, \tag{57}$$

Two values of y_{2n-1} differing by unity must lead to values of Y' differing by q_{2n-1} and in general the converse is true. For, if the expansion in (57) has coefficients satisfying inequalities (14), these inequalities will still be satisfied when y_{2n-1} is increased by one, unless either $y_{2n-1} = a_{2n}$ or $y_{2n-1} = 0$, $y_{2n} = a_{2n+1}$. The latter case is not possible when Y' is a root of (51). And hence $L = q_{2n-1}$ or q_{2n} according as 0, 1 are possible values of y_{2n-1}, or not, when the other coefficients remain unchanged.

When $$y_{2n-1} = 0, \quad -\sigma_{2n-1} > 0.$$

If, by making σ_{2n-1} smaller, we can make it possible that $y_{2n-1} = 0, 1$,

$$b''_{2n-1} w_{2n-1} - \rho_{2n-1} > w_{2n-1} - \sigma_{2n-1}.$$

That is, $$b''_{2n-1} > 1 \quad \text{or} \quad b''_{2n-1} = 1, \quad -\rho_{2n-1} > 0.$$

Both conditions are included if we write

$$\beta'' > w_{2n-1}.$$

Hence the theorem that, assuming inequalities (55) satisfied,

$$L = q_{2n-1} \quad \text{or} \quad q_{2n}$$

according as $\beta'' >$ or $\overline{\gtrless}\, w_{2n-1}$. In the example already considered

$$w_0 > 1 - w_1 > \frac{1}{2} > w_1$$

and therefore $L = q_1 = 1$.

HAVERFORD COLLEGE,
HAVERFORD, PA.

THE THEORY OF SHADOW RAILS

By W. H. Jackson

1. Introduction. It is a common experience to view two sets of railings superposed: when this occurs there are generally to be observed darker vertical bands forming *shadow* rails on a larger scale. The theory of the formation of these bands is the subject of the present paper.

Perhaps, however, the chief interest of the investigation lies in the fact that the discussion enables us to form a very simple model of the phenomena of *group velocity*, observed whenever one train of waves is superposed upon another train of different wave length travelling with a different velocity. The characteristic equation, (20), is obtained in Section 6.

The theory will be simplified by supposing that the rails have no appreciable thickness, only length and breadth, and that they are uniformly spaced, all the rails of one set being in the same plane.

There are no figures illustrating the present paper because it is so easy actually to produce the phenomena described and to do this is so much more useful than to rely upon drawings. Cut out alternate rows of squares from two superposed sheets of squared paper, cutting out less rather than more. If the sheets be separated and held some little distance apart and a dark ground be viewed through the slits made as described, all the phenomena to be discussed can readily be observed.

Let us suppose that the two sets of uniform rails are in parallel planes. Those in the farther plane may be replaced, so far as this phenomenon is concerned, by uniformly spaced rails in the nearer plane. In a given space L let there be n rails of the first set, n' rails of the second set and N shadow rails. By analogy with the theory of beats in sound, we see at once that

$$N = n' - n. \qquad (1)$$

For suppose that $n' > n$, so that the distance between the centres of two consecutive rails of the first set is greater than that between two consecutive rails of the second set. The shadow spaces correspond to places where the two sets of rails have their centers nearly coincident and the shadow rails correspond to places where the rails of one set block the spaces of the other set.

Thus in the distance between the centres of two consecutive shadow spaces, there will be just one more of the second set of rails than of the first set. If in any distance large compared with the distance between consecutive shadow spaces there are N shadow rails, this must be the excess of rails of the second set over rails of the first set. But this is equation (1) put into words.

This intuitive result leads us to expect the average spacing of the shadow rails to be entirely independent of the breadths of the two sets of rails: a result which a more detailed examination confirms.

The following phenomena may easily be observed. As the observer walks past two parallel sets of rails of the same size the shadows travel along with him. If the two sets are not both vertical the slope of the more distant set relative to the nearer set will be exaggerated. But if the nearer set is much more closely spaced than the other both these results are reversed.

If two sets of rails meet at a corner, the shadow rails appear to converge towards or diverge from one special spot according to the direction of motion of the observer.

Finally, if the shadows of a set of rails cast on level ground by the sun be viewed through the rails the shadow bands are found to be curved instead of straight.

The only reference to the subject which the writer has seen is an example given in Chrystal's Algebra,* where the term *ghosts* is applied to places at which the rails appear crowded together. But in that case the rails do not overlap and the shadow rails here considered would not be visible.

Although the theory of the present paper is limited to that case in which uniformly spaced rails produce uniformly spaced shadow rails it is not difficult to trace roughly what would be the changes introduced if the cross-section of the rails instead of being a mathematical line were a rectangle or a circle. In these cases as the eye travels outwards towards the more distant portions of the railing, a point will soon be reached at which it becomes totally opaque. If the rails be replaced by equivalent flat rails, the breadth of these equivalent rails will increase as the eye travels outwards, until the spaces entirely disappear.

It was the study of the phenomena here discussed which led to the theory of integral multiples whose residues, mod 1, lie within given limits, which forms the subject of the preceding paper.

---

* Part II, Chapter XXXII, Exercises XXX, no. 7.

2. Notation. Let $2a$ be the distance between the centers of consecutive rails of the first set and let $2b$ be the breadth of each rail. Let $2a'$, $2b'$ denote corresponding magnitudes for the second set, when viewed in the plane of the first, and suppose that a is greater than a'.

The condition that any space of the first set is entirely closed by a rail of the second set is that c, the distance between their centers, must be less than $b' - (a - b)$. Two cases therefore arise according as b' is greater or less than $a - b$. Only in the former case is it possible for shadow rails to be observed. In the second case, nothing but ghosts, places which are darker because the spaces between consecutive rails are narrowed but not obscured, will be visible.

Let c_r denote the distance of the rth rail-center of the second set from the nearest space-center of the first set and let the direction in which the counting proceeds be the positive direction. Then

$$c_{r+1} = c_r - 2(a - a'), \qquad \text{provided that} \qquad |c_{r+1}| < a. \tag{2}$$

The co-existence of two consecutive values of c, each numerically less than a', implies that

$$a - a' < a' \qquad \text{and therefore} \qquad a' > \tfrac{1}{2}a.^* \tag{3}$$

More generally, the problem to be solved is to find for what values of n, n' the following inequalities are satisfied.

$$b' + b - a \geqq c + 2n'a' - 2na \geqq -(b' + b - a), \tag{4}$$

where $a' > c > -a'$, $a > b$, $a > a' > b'$.

The values of n satisfying (4) determine the numbers of the spaces of the first set which are completely blocked by rails of the second set.

If we write now

$$2a'\beta' = c - (b' + b - a), \quad 2a'\beta = c + (b' + b - a), \; d = a/a', \tag{5}$$

inequalities (4) may be written

$$n' + \beta \;\geqq\; nd \;\geqq\; n' + \beta', \tag{6}$$

n, n' being integers, and the problem is therefore reduced to that discussed in the preceding paper.

---

* If $\frac{1}{3}a < a' < \frac{1}{2}a$, we should write $c_{r+1} = c_r - 2(a - 2a')$.

But in that paper a detailed investigation was made whereas now it is only the simpler average phenomena which come under discussion.

The general condition that shadow rails should exist can readily be written down as a consequence of the result proved at the end of the preceding paper (p. 140). In the notation there used we must find the condition that $L = 1$. A *sufficient* condition is that

$$\beta - \beta' = \frac{b' + b - a}{a'} > w_1 \,. \tag{7}$$

But this is not necessary if $q_2 = a_2 = 1$. In this case the required necessary and sufficient condition is that

$$b' + b - a > a'w_2. \tag{8}$$

3. The breadth of the shadow rails. The boundaries of a shadow rail are determined by rails of the second set which we shall distinguish by the numbers 0 and n, and which satisfy the following inequalities.

$$c_0 > b' + b - a \geqq c_1 > c_2 \cdots > c_{n-1} \geqq -(b' + b - a) > c_n. \tag{9}$$

THEOREM I. *The boundaries of the shadow rails are formed by rails of the more closely spaced set.*

It follows from (9) that c_0 is positive, c_n negative. But c_0, c_n refers to rails of the second set which just fail to cover the nearest space of the first set and hence the open spaces occur on the negative and positive sides in the two cases respectively. It was previously postulated that the direction of counting should be the positive direction and therefore the open spaces which adjoin the rails denoted by the suffixes 0, n lie on the side *away* from the shadow rail. The boundaries of the shadow rail are therefore the boundaries of these two rails which belong to the second set.

THEOREM II. *The breadth of any shadow rail differs from*

$$2\,\frac{b/a + b/a' - a'/a}{1/a' - 1/a}$$

by less than $2a'$.

It follows from inequalities (9) and equations (2) that

$$c_0 > b' + b - a \geqq c_0 - 2(a - a'),$$

$$c_0 - 2(n-1)(a - a') \geqq -(b' + b - a) > c_0 - 2n(a - a').$$

From this it follows that

$$2(a - a') \geqq c_0 - (b' + b - a) > 0, \tag{10}$$

$$2(a - a') \geqq 2n(a - a') - c_0 - (b' + b - a) > 0. \tag{11}$$

Therefore $n - 1$ is the greatest integer contained in

$$\frac{c_0 + b' + b - a}{2(a - a')}$$

and, by addition of the two sets of inequalities (10) and (11), n never differs from

$$\frac{b' + b - a}{a - a'} + 1, = \frac{b' + b - a'}{a - a'},$$

by more than unity.

But the breadth of the shadow rail is $2na' + 2b'$, which therefore differs from

$$2\,\frac{a'b + ab' - a'^2}{a - a'}$$

by less than $2a'$, and this is equivalent to the statement made above.

Corollary. It can be shown by the method of the preceding paper that the average breadth of a large number of consecutive shadow rails approaches $2B$ as a limit, where

$$B = \frac{b/a + b'/a' - a'/a}{1/a' - 1/a}. \tag{12}$$

4. The spacing of the shadow rails.

THEOREM III. *The average space between the corresponding boundaries of a large number of consecutive shadow rails approaches $2A$ as a limit, where*

$$\frac{1}{A} = \frac{1}{a'} - \frac{1}{a}.$$

With reference to the shadow rail next to the one already considered we may write, corresponding to inequalities (10),

$$2(a - a') \geqq c_0 - 2r(a - a') + 2a - (b' + b - a) > 0. \tag{13}$$

From (13) it follows that r is the greatest integer less than

$$\frac{c_0 + 2a - (b' + b - a)}{2(a - a')}$$

and from (10) and (13) together it follows that

$$2(a - a') > 2a - 2r(a - a') > -2(a - a'). \tag{14}$$

Similarly, if the first rail in the Nth shadow rail after the one first considered is denoted by the suffix r_N, we have corresponding to inequalities (14)

$$1 > \frac{Na}{a - a'} - r_N > -1,$$

from which it follows that

$$\lim_{N=\infty}\left(\frac{r_N}{N}\right) = \frac{a}{a - a'}.$$

Therefore

$$A = \lim_{N=\infty}\left(\frac{r_N a'}{N}\right) = \frac{aa'}{a - a'},$$

which is equivalent to the result stated above.

Corollary. If n, n', N have the meanings used in equation (1), Theorem III may be stated in the form

$$\lim_{N=\infty}\left(\frac{n' - n}{N}\right) = 1. \tag{15}$$

Similarly the corollary to Theorem II may be stated in the form

$$\lim_{N=\infty}\left(\frac{2nb + 2n'b' - 2NB}{L}\right) = \lim_{N=\infty}\left(\frac{n}{n'}\right). \tag{16}$$

That is, in any given distance the total space covered by the rails of the two sets taken separately exceeds the space covered by the shadow rails by an amount which depends only on n, n' and *is independent of the breadths* of the two sets of rails.

5. Ghosts. If we seek to extend the results of the two last sections to ghosts, we must no longer compare the *beginnings* of two consecutive bands because in this case the shadow has no definite limits. But we can

pick out in each ghost one rail of the second set for which

$$\tfrac{1}{2}(a - a') \geqq c > -\tfrac{1}{2}(a - a') \tag{17}$$

and call this the center rail.

Theorem IV. *The average space between center rails of a large number of consecutive ghosts approaches $2A$ as a limit, where*

$$\frac{1}{A} = \frac{1}{a'} - \frac{1}{a}.$$

With reference to the Nth ghost beyond the one already considered

$$\tfrac{1}{2}(a - a') \geqq c - 2r_N(a - a') + 2Na > -\tfrac{1}{2}(a - a'). \tag{18}$$

From inequalities (17) and (18)

$$(a - a') > 2Na - 2r_N(a - a') > -(a - a'),$$

whence it follows that

$$\frac{1}{2} > \frac{Na}{a - a'} - r_N > -\frac{1}{2}.$$

From this point the proof proceeds exactly as in Theorem III.

6. The effect of motion. The edge of a shadow rail always coincides with that of a rail of the second set. Hence if the second set moves with velocity v relative to the first set, the relative velocity of the edge of the shadow rail is in general equal to v but an instantaneous jump occurs whenever this shadow edge changes from one rail to the next resulting in a displacement $2a'$. The space-time graph for the shadow edge is therefore made up of a series of steps. First a distance $2(a - a')$ at uniform velocity v, next an instantaneous jump of $2a'$, giving an average velocity V relative to the first set where

$$V = \frac{av}{a - a'}. \tag{19}$$

In the application to group velocity suppose two sets of waves superposed of wave lengths λ, $\lambda + \delta\lambda$ and travelling with velocities v, $v + \delta v$ respectively. It follows from (19) that the group velocity is

$$V = v - \lambda\frac{dv}{d\lambda}. \tag{20}$$

Suppose an observer to walk with velocity v_0 parallel to two sets of rails and at a distance from them of d and d' respectively. Let $2a$, $2a'$ be the distances between the centers of consecutive rails in the two sets respectively. Viewed in the plane of the nearer set, the second set are spaced at distances $2a'd/d'$ and move forward with velocity

$$v = \frac{d' - d}{d'} v_0 . \tag{21}$$

If

$$a'd < ad'$$

the shadow rails move *forward* with velocity

$$V = \frac{a(d' - d)}{ad' - a'd} v_0 = v_0 + \frac{(a' - a)d}{ad' - a'd} v_0. \tag{22}$$

They always move in the same direction as the observer, but overtake him or fall behind according as the more distant set is more widely spaced than the nearer set or not.

If

$$a'd > ad',$$

the shadow rails are bounded by rails of the nearer set and move backwards relatively to the further set with velocity

$$\frac{(d' - d)a'd}{(a'd - ad')d'} v_0 .$$

The velocity relative to the rails of the *nearer* set is given, as in the first case, by equations (22).

Now let the observer approach from a great distance so that the ratio d/d' decreases from unity to zero. Two cases arise when

$$b' + b > a, \qquad \text{according as} \qquad a \gtrless a'. \tag{23}$$

In the first case it follows, by replacing a', b' in Theorem II by $a'd/d'$, $b'd/d'$ respectively, that the size of the shadow rails, viewed in the plane of the nearer set, continuously decreases as the observer approaches. Ghosts, as defined in section 2, will be observed when

$$\frac{d}{d'} < \frac{a - b}{b'} . \tag{24}$$

From Theorems III and IV it may be seen that the spacing of the shadow rails or ghosts, as the case may be, continually decreases whilst the ratio B/A increases towards the limit $(b/a + b'/a')$.

In the second case defined by (23) the shadow rails become infinite when $a'd = ad'$. From this point on, the effect of approaching the rails is much as in the first case.

7. The effect of angular inclination. Let the slope, the tangent of the angle of inclination, of the second set of rails relative to the first set be s and let S denote the slope of the shadow rails relative to the first set. The open spaces common to both sets are parallelograms arranged in rows. When s is small the parallelograms are long.

The results to be obtained in this section follow easily from those at the beginning of the previous section. The figure to be considered may be regarded as the space-time graph of the motion of the second set of rails relative to the first with uniform velocity s if the edge of a rail of the first set is taken as the time axis. Applying equation (19), the average slope of the shadow rails relative to the first set is given by

$$S = \frac{a}{a - a'} s. \tag{25}$$

8. The effect of perspective. When the planes of the two sets of rails are not parallel, the projections of the second set on the plane of the first are concurrent instead of parallel. As in the case just considered, the boundaries of the shadow bands may be smoothed out by a line joining the common points of a rail 1 of the first set and rail 1 of the second set, rail 2 of the first set and rail 2 of the second set, and so on. But in this case the line is not straight, but curved.

Take as axis of x the vanishing line of the plane of the second set when the plane of the first set is the picture plane. Take as axis of y the straight line through the vanishing point of the second set of rails, parallel to the first set. The equations of the edges of the nth rail of the first set and the n'th rail of the second set may be written respectively

$$x = 2na + c \pm b, \tag{26}$$

$$x = y(2n'a' + c' \pm b'). \tag{27}$$

The shadow bands will be more clearly defined in those parts of the plane in

which the angle between the two sets of rails is small. Such a region is that included between

$$x = \pm \alpha\, y, \tag{28}$$

where α has some value such as $\frac{1}{2}$ or $\frac{1}{3}$.

The boundaries of the shadow bands may now be defined by writing

$$n' - n = N \tag{29}$$

and eliminating n, n' from equations (26), (27) and (29). The equation thus obtained is

$$\frac{x}{a'y} - \frac{x}{a} = 2N + C, \tag{30}$$

where

$$C = \frac{c' \pm b'}{a'} - \frac{c \pm b}{a}\,. \tag{31}$$

Equation (30) represents a hyperbola with asymptotes

$$a'y = a,\ x + (2N + C)\, a = 0. \tag{32}$$

The shadow bands are therefore bounded by hyperbolas with asymptotes parallel to the first set of rails and the vanishing line of the plane of the second set of rails respectively. The latter asymptote is fixed and therefore common to all the bands. It is that line for which the distance between intersections with the center lines of consecutive rails is the same for both sets.

An important exception occurs when both the two sets of rails are parallel to the vanishing line. In this case the shadow bands must reduce to shadow rails parallel to the given sets but they will no longer be uniformly spaced. For our present purpose it is sufficient to consider the case in which the observer is at a distance large compared with the shadow rails. In this case the spacing of the shadow rails in the neighborhood of any point may be calculated as if the second set appeared as uniform rails when viewed in the plane of the first set. Let the planes of the two sets of rails make angles θ, θ' with the plane passing through their join and the observer's eye. Let a plane through a rail to be considered and the observer's eye make an angle ϕ with this plane.

The distance apart of rails in the second set being $2a'$ and their apparent distance, viewed in the plane of the first set, at this point being $2a''$,

$$a'' = a' \frac{\sin\theta \sin^2(\phi + \theta')}{\sin\theta' \sin^2(\phi + \theta)}. \tag{33}$$

If $\theta' > \theta$, as ϕ increases from 0 to $\pi - \theta'$, a'' decreases from $a' \sin\theta'/\sin\theta$ to 0. Two cases therefore arise according as

$$a' \sin\theta' \gtrless a \sin\theta. \tag{34}$$

In the first case A increases with ϕ and becomes infinite when

$$a \sin\theta' \sin^2(\phi + \theta) = a' \sin\theta \sin^2(\phi + \theta'). \tag{35}$$

For greater values of ϕ, A continually decreases, as also occurs in the second of the two cases indicated by (34).

Since, by Theorem I, the direction of motion of the shadow rails relative to the larger set is the same as that of the more closely spaced set, and since in the neighborhood of the value of ϕ determined by equation (35) A is large, it follows that the effect of motion on the part of the observer in any direction but that of ϕ will be to cause the shadow rails on opposite sides of this critical position to move in opposite directions.

Further, the rails appear to converge towards or diverge from this spot according as the observer's path makes an angle greater or less than ϕ with the plane through his eye and the join of the planes of the two sets of rails.

If the direction of motion is towards the rails, the shadows will converge towards or diverge from the point approached because the relative motion of the two sets of rails changes sign at this point. But this motion is slow, because of the small relative velocity at this point, while that just considered is fast, because of the large magnification.

Finally, equation (35) may be put into the more convenient form

$$\frac{\cot\phi + \cot\theta}{\cot\phi + \cot\theta'} = \left(\frac{a'\sin\theta'}{a\sin\theta}\right)^{\frac{1}{2}}. \tag{36}$$

If we put

$$a = a', \qquad \theta' = \theta + \tfrac{1}{2}\pi, \qquad x = \cot\phi, \qquad c^2 = \cot\theta,$$

we obtain

$$x = \frac{c^3 + 1}{c^2 - c}.$$

The value of x is positive only when $c > 1$, that is when $\theta < 45°$. A short calculation shows that the minimum value of x, and therefore the maximum value of ϕ, occurs when c lies near 2·2, making θ and ϕ each about 12°.

The numerical values have been chosen to represent the case of two uniform sets of rails meeting at right angles. Observation confirms these results, though this is a case in which the substitution of round or square rails for flat ones makes a considerable difference. Owing to the large magnification at the critical angle, any slight deviation of the plane of either set of rails from the vertical causes an appreciable curvature in the shadow rails.

Haverford College,
Haverford, Pa.

ON A NEW METHOD OF COMPUTING THE ROOTS OF BESSEL'S FUNCTIONS

By William Marshall

The object of this paper is to show the application of a certain method to the computation of the roots of Bessel's functions. This method which was employed by me in my Doctor's Thesis* in the computation of the roots of the functions of the elliptic cylinder, and which is generally applicable in similar problems is essentially this: The coefficients of certain asymptotic expansions which are necessary in the computation, are not found by arithmetic processes, but are determined from recurrence formulas which arise when certain differential equations are formally satisfied. The results are of course not new. The roots of the Bessel's functions of order zero as determined from the semiconvergent expansion were first given by Stokes.† The roots of the Bessel's functions of order n are given by McMahan‡ and are found in Gray and Mathews.§

The ordinary Bessel's equation, namely

$$(1) \qquad \frac{d^2J_n(x)}{dx^2} + \frac{1}{x}\frac{dJ_n(x)}{dx} + \left(1 - \frac{n^2}{x^2}\right) J_n(x) = 0,$$

can by means of the substitution $U = X^{\frac{1}{2}}J_n(x)$ be reduced to the form

$$(2) \qquad \frac{d^2U}{dx^2} + \left(1 - \frac{n^2 - \frac{1}{4}}{x^2}\right) U = 0.$$

A solution of this in the neighborhood of $x = \infty$ is

$$(3) \qquad U = C \{ P \cos (a - x) + Q \sin (a - x) \},$$

* Diss. Zürich, 1909, *American Journal of Mathematics*, vol. 31, no. 4, Oct., 1909.

† G. G Stokes, *Trans. Cambridge Phil. Soc.*, vol. 9, part I, or *Math. and Phys. Papers*, vol. 2, p. 329.

‡ Annals of Mathematics vol. 9, Jan., 1895.

§ *Treatise on Bessel's Functions*, p. 241.

where C and a are integration constants* and where P and Q represent the following asymptotic series:

$$(4)\quad \begin{cases} P = 1 - \dfrac{(4n^2-1)(4n^2-9)}{1.2(8x)^2} \\ \qquad\qquad + \dfrac{(4n^2-1)(4n^2-9)(4n^2-25)(4n^2-49)}{1.2.3.4\,(8x)^4} - \cdots \\ Q = \dfrac{4n^2-1}{8x} - \dfrac{(4n^2-1)(4n^2-9)(4n^2-25)}{1.2.3\,(8x)^3} + \cdots \end{cases}$$

If $C = \sqrt{\dfrac{2}{\pi}}$ and $a = \dfrac{2n+1}{4}\pi$, and if M and ψ are determined so that

$$(4')\qquad M\cos\psi = P \text{ and } M\sin\psi = Q,$$

then from (3)

$$(5)\qquad U = CM\cos(a - x - \psi),$$

and as $U = x^{\frac{1}{2}}J_n(x)$,

$$J_n(x) = \sqrt{\frac{2}{\pi x}}\, M\cos\left(\frac{2n+1}{4} - x - \psi\right),$$

and the roots of $J_n(x) = 0$ can be calculated from the relation

$$(6)\qquad \frac{2n+1}{4}\pi - x - \psi = -\frac{2k-1}{2}\pi(k = 1, 2, \cdots).$$

M and ψ have been calculated approximately and with great labor from the relations $M = \sqrt{P^2 + Q^2}$, $\psi = \tan^{-1}(Q/P)$ respectively. They may be obtained much more simply by the aid of equation (2) as follows. $U = CM\cos(a - x - \psi)$ will satisfy (2) whatever values may be given to C and a. Let $C = 1$, and $a = 0$. Then a solution of (2) is

$$(7)\qquad U = M\cos(\psi + x),$$

and M and ψ are to be determined so that (2) is satisfied. From (7) we obtain

$$(8)\qquad U' = \cos(\psi + x)M' - M(\psi' + 1)\sin(\psi + x),$$

* Gray and Mathews, l. c., p. 37.

$$(9)\qquad U'' = \cos(\psi + x)\, M'' - 2M' \sin(\psi + x)(\psi' + 1) - M\cos(\psi + x)(\psi' + 1)^2 - M\sin(\psi + x)\psi''.$$

Substituting the values of U'' and U in (2) we obtain

$$(10)\quad M''\cos(\psi + x) - 2M'\sin(\psi + x)(\psi' + 1) - M\cos(\psi + x)(\psi' + 1)^2 - M\sin(\psi + x)\psi'' + M\cos(\psi + x)\left(1 - \frac{n^2 - \frac{1}{4}}{x^2}\right) = 0.$$

Since the coefficients of $\sin(\psi + x)$ and $\cos(\psi + x)$ must separately vanish, we obtain the two following differential equations for M and ψ, namely,

$$(11)\qquad 2M'(\psi' + 1) + M\psi'' = 0,$$

$$(12)\qquad M'' - M(\psi' + 1)^2 + M\left(1 - \frac{n^2 - \frac{1}{4}}{x^2}\right) = 0.$$

Equation (11) integrates readily into

$$(13)\qquad \psi' + 1 = \frac{k}{M^2},$$

where k is the constant of integration, which may be determined from the fact that $\psi = \tan^{-1}\frac{Q}{P}$, whence $\psi' = \frac{PQ' - P'Q}{P^2 + Q^2}$. Therefore $k = 1$ and we have

$$(14)\qquad \psi' + 1 = \frac{1}{M^2}.$$

Substituting this value of ψ' in (12) we have a differential equation for M, namely

$$(15)\qquad M'' - \frac{1}{M^3} + M\left(1 - \frac{n^2 - \frac{1}{4}}{x^2}\right) = 0.$$

Now from (4′) we see that the first term of M must be 1; we therefore assume as a solution of (15)

$$(16)\qquad M = 1 + \frac{A_2}{x^2} + \frac{A_4}{x^4} + \frac{A_6}{x^6} + \cdots,$$

where the A's are undetermined constants. From (16) we obtain

$$(17)\qquad M'' = \frac{3.2\, A_2}{x^4} + \frac{5.4\, A_4}{x^6} + \frac{7.6 A_6}{x^8} + \cdots.$$

Also, by the Binomial Theorem,

$$\frac{1}{M^3} = 1 + \frac{a_2}{x^2} + \frac{a_4}{x^4} + \frac{a_6}{x^6} + \cdots, \tag{18}$$

where

$$\begin{aligned}
a_2 &= -3A_2, \\
a_4 &= \frac{-4A_2 a_2}{2} - 3A_4, \\
a_6 &= \frac{-5A_2a_4 - 7A_4a_2}{3} - 3A_6, \\
a_8 &= \frac{-6A_2a_6 - 8A_4a_4 - 10A_6a_2}{4} - 3A_8, \\
&\cdots\cdots\cdots\cdots
\end{aligned} \tag{19}$$

or in terms of the A's alone

$$\begin{aligned}
a_2 &= -3A_2, \\
a_4 &= 6A_2^2 - 3A_4, \\
a_6 &= -10A_2^3 + 12A_2A_4 - 3A_6, \\
a_8 &= 15A_2^4 - 30A_2^2A_4 + 12A_2A_6 + 6A_4^2 - 3A_8, \\
&\cdots\cdots\cdots\cdots
\end{aligned} \tag{20}$$

Substituting now in (15) the value of M from (16), M'' from (17), and $1/M^3$ from (18), we have

$$\begin{aligned}
&\frac{3.2\,A_2}{x^4} + \frac{5.4\,A_4}{x^6} + \frac{7.6\,A_6}{x^8} + \cdots \\
&+ \frac{A_2}{x^2} + \frac{A_4}{x^4} + \frac{A_6}{x^6} + \frac{A_8}{x^8} + \cdots \\
&- \frac{n^2 - \frac14}{x^2} - \frac{(n^2 - \frac14)A_2}{x^4} - \frac{)n^2 - \frac14)A_4}{x^6} - \frac{(n^2 - \frac14)}{8^8}A_6 - \cdots \\
&- \frac{a_2}{x^2} - \frac{a_4}{x^4} - \frac{a_6}{x^6} - \frac{a_8}{x^8} - \cdots = 0.
\end{aligned} \tag{21}$$

Equating to zero the coefficients of like powers of x and making use of the relations (19) we have

$$
\begin{aligned}
4A_2 &= n^2 - \tfrac{1}{4}, \\
4A_4 &= (n^2 - \tfrac{1}{4} - 3.2)A_2 - \frac{4A_2a_2}{2}, \\
4A_6 &= (n^2 - \tfrac{1}{4} - 5.4)A_4 - \frac{5A_2a_4 + 7A_4a_2}{3}, \qquad (22) \\
4A_8 &= (n^2 - \tfrac{1}{4} - 7.6)A_6 - \frac{6A_2a_6 + 8A_4a_4 + 10A_6a_2}{4}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

Formulas (22) then enable us to compute the coefficients in the expansion of $M = \sqrt{P^2 + Q^2}$ without serious labor. These formulas are, in terms of the A's,

$$
\begin{aligned}
4A_2 &= n^2 - \frac{1}{4}, \\
4A_4 &= \left(n^2 - \frac{25}{4}\right)A_2 + 6A_2^2, \\
4A_6 &= \left(n^2 - \frac{81}{4}\right)A_4 - 10A_2^3 + 12A_2A_4, \qquad (23) \\
4A_8 &= \left(n^2 - \frac{169}{4}\right)A_6 + 15A_2^4 - 30A_2^2A_4 + 12A_2A_6 + 6A_4^2, \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

or in terms of n alone,

$$
\begin{aligned}
A_2 &= \frac{1}{2^4}(4n^2 - 1), \\
A_4 &= \frac{1}{2^9}(4n^2 - 1)(20n^2 - 53), \\
A_6 &= \frac{1}{2^{13}}(4n^2 - 1)(240n^4 - 2488n^2 + 4477), \qquad (24) \\
A_8 &= \frac{1}{2^{19}}(4n^2 - 1)(12480n^6 - 322192n^4 + 2023460n^2 - 3066403), \\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

or, if we set for shortness $4n^2 = m$,

$$A_2 = \frac{1}{2^4}(m-1),$$

$$A_4 = \frac{1}{2^9}(m-1)(5m-53),$$

$$(25) \qquad A_6 = \frac{1}{2^{13}}(m-1)(15m^2 - 622m + 4447),$$

$$A_8 = \frac{1}{2^{19}}(m-1)(195\,m^3 - 20137m^2 + 505865m - 3066403),$$

. . . .

From these coefficients in the expansion of M, we may readily compute those in the development of $\psi = \tan^{-1}Q/P$. We have from (14)

$$(26) \qquad \psi' = \frac{1}{M^2} - 1.$$

Then ψ' is of the form

$$(27) \qquad \psi' = \frac{c_2}{x^2} + \frac{c_4}{x^4} + \frac{c_6}{x^6} + \cdots$$

Integrating, and setting the constant of integration equal to zero, since $\tan\psi = Q/P$, we have

$$(28) \qquad \psi = \frac{-C_2}{x} - \frac{C_4}{3x^3} - \frac{C_6}{5x^5} - \frac{C_8}{7x^7} - \cdots,$$

where the C's are simply the coefficients in the expansion of $1/M^2$ by the Binomial theorem, that is,

$$C_2 = -2A_2,$$

$$2C_4 = -3A_2C_2 - 4A_4,$$

$$(29) \qquad 3C_6 = -4A_2C_4 - 5A_4C_2 - 6A_6,$$

$$4C_8 = -5A_2C_6 - 6A_4C_4 - 7A_6C_2 - 8A_8,$$

. . . .

or, in terms of the A's,

$$
(30)\qquad
\begin{aligned}
C_2 &= -\,2A_2,\\
C_4 &= \quad 3A_2^2 - 2A_4,\\
C_6 &= -\,4A_2^3 + 6A_2A_4 - 2A_6,\\
C_8 &= \quad 5A_2^4 - 12A_2^2A_4 + 6A_2A_6 + 3A_4^2 - 2A_8,\\
&\quad \cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

These coefficients in the expansion of ψ are used in the computation of the roots of $J_n(x)$ in the following manner: an asymptotic value of $J_n(x)$ is *

$$
(31)\qquad J_n(x) = \sqrt{\frac{2}{\pi x}}\left[P\cos\left(\frac{2n+1}{4}\pi - x\right) + Q\sin\left(\frac{2n+1}{4}\pi - x\right)\right].
$$

This may be written

$$
(32)\qquad J_n(x) = \sqrt{\frac{2}{\pi x}}\,M\cos\left(\frac{2n+1}{4}\pi - x - \psi\right),
$$

which vanishes when

$$
(33)\qquad \frac{2n+1}{4}\pi - x - \psi = -\frac{2k-1}{2}\pi,
$$

that is, when

$$
(34)\qquad x = \frac{\pi}{4}(2n - 1 + 4k) - \psi,
$$

where k is any integer.

If we put for shortness $\frac{\pi}{4}(2n - 1 + 4k) = \beta$,

then the values of x for which $J_n(x)$ vanishes are found by solving the equation

$$
(35)\qquad x = \beta - \psi,
$$

or

$$
(36)\qquad x = \beta + \frac{C_2}{x} + \frac{C_4}{3x^3} + \frac{C_6}{5x^5} + \cdots.
$$

Equation (36) is best solved by the method of successive approximations. The successive approximations are

* Gray and Mathews, *Treatise on Bessel's Functions*, p. 40.

$$
(37)\begin{cases}
x_1 = \beta_1, \\
x_2 = \beta + \dfrac{C_2}{\beta}, \\
x_3 = \beta + \dfrac{C_2}{\beta} + \dfrac{\dfrac{C_4}{3} - C_2^2}{\beta^3}, \\
x_4 = \beta + \dfrac{C_2}{\beta} + \dfrac{\dfrac{C_4}{3} - C_2^2}{\beta^3} + \dfrac{2C_2^3 - \dfrac{4}{3}C_2C_4 + \dfrac{C_6}{5}}{\beta^5}, \\
x_5 = \beta + \dfrac{C_2}{\beta} + \dfrac{\dfrac{C_4}{3} - C_2^2}{\beta^3} + \dfrac{\dfrac{C_6}{5} - \dfrac{4}{3}C_2C_4 + 2C_2^3}{\beta^5} \\
\qquad + \dfrac{\dfrac{C_8}{7} - 5C_2^4 - \dfrac{6}{5}C_2C_6 + 5C_2^2C_4 - \dfrac{C_4^2}{3}}{\beta^7}.
\end{cases}
$$

Expressing these coefficients of β^{-1}, β^{-3} . . . in terms of m by means of the relations (25) and (30) we have for the kth root of $J_n(x)$,

$$
(38) \qquad x_k = \beta - \frac{m-1}{2^3\beta} - \frac{(m-1)}{3\cdot 2^7}\cdot\frac{(7m-31)}{\beta^3} - \frac{(m-1)}{15.2^{10}}\cdot\frac{(83m^2 - 982m + 3779)}{\beta^5}
$$

$$
- \frac{(m-1)}{105.2^{15}}\cdot\frac{(6949m^3 - 153855m^2 + 1585743m - 6277237)}{\beta^7} - \cdots
$$

where $\qquad \beta = \frac{\pi}{4}(2n + 4k - 1),$

and $\qquad m = 4n^2.$

This result agrees with that given by McMahan;* there is however a slight error in the formula as given by Gray and Mathews,† due to the dropping of a figure in the coefficient of β^{-7}.

Purdue University.
October, 1909.

---

* Annals of Mathematics, ser. 2, vol. 9, p. 25.
† Loc. cit., p. 241.

A FUNCTIONAL EQUATION FOR THE SINE

By Edward B. Van Vleck *

In 1821 Cauchy † proved that the only real continuous solutions of the equation

$$\phi(x + y) + \phi(x - y) = 2\,\phi(x)\,\phi(y),$$

other than the trivial solutions $\phi(x) \equiv 0$ or 1, are the functions $\phi(x) = \cos ax$ and $\phi(x) = \cosh ax$. As analytic functions these two solutions are one and the same, but considered as real functions of a real variable they are distinct. The object of this note is to give a functional equation satisfied uniquely by the cosine, or, what amounts to the same thing, one satisfied uniquely by the sine. It appears to me much the simplest functional definition yet given, ‡ and the fundamental properties of the solution follow from the equation with very great rapidity. The only assumption concerning its character is that it is real and continuous.

To define the sine I use the equation

$$(I) \qquad f(x - y + A) - f(x + y + A) = 2f(x)f(y),$$

in which A is a real constant. The following properties immediately result.

1) Put $x = y = 0$. Then $f(0) = 0$.

2) Replace y by $-y$. The left hand member of (I) changes sign, and therefore $f(y) = f(-y)$. In other words, $f(x)$ is an odd function.

3) Exchange x and y. It follows that

$$f(x - y + A) = f(y - x + A) = -f(x - y - A).$$

Hence the addition of $2A$ to the argument changes the sign of the function, and therefore $4A$ is a period of $f(x)$. Without loss of generality we may henceforth suppose A to be positive.

---

* Presented to the American Mathematical Society (Chicago).

† *Analyse Algebrique*, p. 114; *Oeuvres* ser. 2, vol. 3, p. 106.

‡ For other functional definitions see Moore, Annals of Mathematics, (1), vol. 9 (1895), p. 43; Lunn, Ibid (2), vol. 10 (1908), p. 37; and Osgood, *Lehrbuch der Funktionentheorie*, vol. 1, p. 510.

4) Replace x by $x + A$ and y by $y + A$. Equation (I) then gives with the aid of 3),

$$(1) \qquad f(x - y + A) + f(x + y + A) = 2f(x + A)\, f(y + A).$$

Put now $\phi(u) \equiv f(u + A)$. Then

$$\phi(x - y) + \phi(x + y) = 2\phi(x)\, \phi(y).$$

Thus $f(x + A)$ satisfies Cauchy's functional equation, and by 3) it must be the periodic solution of that equation. Selecting this solution, we have from Cauchy's work

$$f(x) = \cos \frac{\pi}{2A}(x - A) = \sin \frac{\pi x}{2A}.$$

The discussion can, however, be carried forward without having recourse to Cauchy's equation. Thus

5) If we put $x = y = A$ in (I), we have

$$f(A) + f(A) = 2f^2(A),$$

whence $f(A) = 0$ or 1. But if $f(A) = 0$, it follows from (I) by putting $y = A$ that $f(x)$ identically vanishes. Putting aside this trivial solution, we must take $f(A) = 1$.

6) Put $y = x + A$. Then from (I),

$$f(2x) = 2f(x)\, f(x + A).$$

This equation was used by Moore in his determination of the sine, but supplementary conditions of equal importance and restrictive scope were necessarily added to obtain a definite function.

7) By placing $x = y$ in both (I) and (1) and adding we obtain

$$(2) \qquad 1 = f^2(x) + f^2(x + A).$$

8) It may be shown next that $f(x)$ is positive and steadily increases when x increases from 0 to A.

For if possible, let us suppose that x vanishes for some value B between these limits. If then we set $y = B$ in (I), we obtain

$$f(x - B + A) = f(x + B + A).$$

Consequently $2B$ must be a period of $f(x)$. There must be a smallest B, since otherwise the function would have a period as small as we please and, being continuous, would therefore be constant. I shall use this smallest B.

For $x = y$, equation (I) becomes

$$1 - f(2x + A) = 2f^2(x).$$

Hence $|f(2x + A)| < 1$ for values of x between 0 and B. It follows then by replacing x by $2x$ in (2) that $|f^2(2x)| > 0$ for the same range of values of x. But this gives a contradiction, inasmuch as $f(2x)$ by our hypothesis must vanish for $x = B/2$. Consequently $f(x)$ can not vanish for $0 < x \leqq A$. Since also $f(A) = 1$, it must be positive.

Furthermore, if we put $x = 0$ in (I) and then replace y by x, we have

$$f(A + x) = f(A - x).$$

This shows that $f(x)$ is also positive for $A < x < 2A$.

To see finally that $f(x)$ steadily increases with x in the interval $(0, A)$, replace x by $x + h + A$ and y by h in (I). The equation becomes

$$(3) \qquad f(x + 2h) - f(x) = 2f(x + h + A) f(h).$$

For positive $h < A$ and for $0 < x + h < A$ the right hand member is positive, and therefore any small positive increment $2h$ in the argument results in an increase of $f(x)$ as long as x lies in the interval $(0, A)$.

9) From (3) we get the difference quotient

$$(4) \qquad Q(x + 2h, x) \equiv \frac{f(x + 2h) - f(x)}{2h} = f(x + h + A)\frac{f(h)}{h}.$$

This shows that *if $f(h)/h$ has a limit c for $h = 0$, $f(x)$ will have a derivative, of which the value is $cf(x + A)$.*

10) A proof of the existence of a limit for $f(h)/h$ when h approaches 0 has been given by Osgood,* and his proof can be based upon the equations

* Loc. cit., pp. 512–513.

already here obtained. Another simple proof, based directly upon a consideration of the difference quotient, is as follows.

Putting $x = 0$ in (4) we see that

$$\frac{f(2h)}{2h} < \frac{f(h)}{h}.$$

If then x approaches 0 over any set of positive values $2h$, h, $h/2$, $h/4$, $\cdots$, the ratio $f(x)/x$ will steadily increase.

We shall next prove that it can not increase beyond all limit. Divide the interval $(0, 2h)$ into 2^n equal parts of length $h/2^{n-1}$. The difference quotient taken for any two consecutive points of division, $p - h/2^{n-1}$ and p, has by (4) the value

$$Q\left(p, p - \frac{h}{2^{n-1}}\right) = f\left(p - \frac{h}{2^n} + A\right) \cdot \frac{f\left(\frac{h}{2^n}\right)}{\frac{h}{2^n}}.$$

Hence since $f(x + A)$ decreases in the x-interval $(0, A)$ we have the inequality

$$f(A + 2h) \cdot \frac{f\left(\frac{h}{2^n}\right)}{\frac{h}{2^n}} < Q\left(p, p - \frac{h}{2^{n-1}}\right) < \frac{f\left(\frac{h}{2^n}\right)}{\frac{h}{2^n}}.$$

Now if (a, b) is any interval of the x-axis and c is an internal point of division, by a very fundamental but simple theorem* regarding the difference quotient, $Q(a, b)$ is intermediate in value between $Q(a, c)$ and $Q(c, b)$. By obvious extension, if the interval is divided into any number of parts by the points $c_0 = a, c_1, c_2, \cdots, c_{n-1}, c_n = b$, the value of $Q(a, b)$ will lie between the greatest and least values of $Q(c_i, c_{i+1})$. Consequently

$$(5) \qquad f(A + 2h) \frac{f\left(\frac{h}{2^n}\right)}{\frac{h}{2^n}} < Q(p, 0) < \frac{f\left(\frac{h}{2^n}\right)}{\frac{h}{2^n}}.$$

* Cf. Baire, *Annali di Matematica*, (3) vol. 3 (1899), p. 103.

In this inequality let n increase indefinitely. If $f(x)/x$ increases without limit when x traverses the values $h/2$, $h/4$, $h/8$, . . . it will follow at once that $Q(p, 0)$ is infinite. As this is impossible, we conclude that $f(x)/x$ for this set of values increases to a finite limit c.

Consider now the dense set, S, consisting of all points of division of the interval $(0, 2h)$ which are obtained by taking successively $n = 1, 2, 3, \cdots$. Let p be any point of the set. By increasing n indefinitely in (5) we obtain

$$(6) \qquad cf(2h + A) \leqq Q(p, 0) \equiv \frac{f(p)}{p} < c.$$

We can contract the interval $2h$ as much as we please by replacing h by $h/2^n$. It follows then directly from (6) that $f(x)/x$ has the limit c when x approaches 0 over the set S which is everywhere dense in the interval $(0, 2h)$.

Now if this is true for a *continuous* $f(x)$, the conclusion can be extended immediately from the dense set S to the interval itself. For if x' is any point of the interval not included in S, a point p of S can be found so near to x' that

$$\frac{f(x')}{x} = \frac{f(p)}{p}(1 + \epsilon),$$

where ϵ is as small as we wish. Consequently $f(x')/x'$ has a common limit with $f(p)/p$. Since $f(x)$ is odd, the limit is the same for negative h.

The essential properties of $f(x)$, including its differentiability, have been established, from which its identification with $\sin cx$ readily follows. This can be done in various ways. Thus, for example, if the derivative

$$\frac{df(x)}{dx} = cf(x + A) \equiv c\sqrt{1 - f^2(x)}$$

is used for this purpose, it follows immediately by integration that $f(x) = \sin cx$.

University of Wisconsin,
Madison, Wis.

PERIODIC DECIMAL FRACTIONS

By W. H. Jackson

When at school the writer was shown that vulgar fractions with unit numerator and a denominator ending in 9 can be expressed as repeating decimals according to the simple rule illustrated in the following example.

Suppose it is required to express $1/39$ as a repeating decimal. The last figure is 1. To obtain the preceding figures

$1 \times 4 \quad\quad = 4,$	put down 4,
$4 \times 4 \quad\quad = 16,$	put down 6, carry 1,
$6 \times 4 + 1 = 25,$	put down 5, carry 2,
$5 \times 4 + 2 = 22,$	put down 2, carry 2,
$2 \times 4 + 2 = 10,$	put down 0.

The result is given below.

$$\frac{1}{39} = \cdot\dot{0}\ 2\ 5\ 6\ 4\ \dot{1}$$

It is at once clear that the process may be worked in the opposite direction.

$$\begin{aligned} 1 &= 1 + 0 \times 4, \\ 10 &= 2 + 2 \times 4, \\ 22 &= 2 + 5 \times 4, \\ 25 &= 1 + 6 \times 4, \\ 16 &= \quad\ \ 4 \times 4, \\ 4 &= \quad\ \ 1 \times 4. \end{aligned}$$

Since that time the writer has found no reference to the rule or to the following theorem, from which the rule follows as a special case.

Theorem. If a be any number prime to 10, and b, less than 10, be chosen so that

$$ab = 10m - 1 \equiv 9, \text{ mod } 10,$$

the number which must be repeated to form the decimal equivalent of $1/a$ may be written

$$b[1 + 10m + (10m)^2 + (10m)^3 + \cdots + (10m)^{s-1}] - k10^s,$$

where s is the exponent to which 10 belongs, mod a, and k is a positive integer.

By way of proof it is only necessary to note the identity

$$\frac{-1 + 10^s}{ab} = \frac{-1 + 10^s}{-1 + 10m} = 1 + 10m + (10m)^2 + \cdots + (10m)^{s-1} - 10^s \frac{m^s - 1}{ab}$$

and to multiply through by b. When this is done the left-hand side is an integer and we can therefore write, as is otherwise evident,

$$\frac{m^s - 1}{a} = k, \text{ some positive integer.}$$

From this equation it follows that

$$\frac{1}{a} = b\left[\frac{1}{10^s} + \frac{m}{10^{s-1}} + \frac{m^2}{10^{s-2}} + \cdots + \frac{m^{s-1}}{10}\right] - k + \frac{1}{a} \cdot \frac{1}{10^s}.$$

When, as in the example above, a ends in a nine, we may choose

$$b = 1, \quad m = \frac{a + 1}{10}.$$

It is easily seen that this formula is equivalent to the rule as stated. The arrival at the end of the process may be recognized by the figures beginning to repeat. Evidently when m is less than 10 this will occur after the application of the rule gives 10 as a product.

The range of numbers for which this method is of practical use is not large, as may be seen from the following table giving values of a corresponding to simple values of m.

m	a	m	a
1	3,9	20	199
2	19	25	249,83
3	29	30	299
4	39,13	40	399,133,57
5	49,7	50	499
6	59	60	599
7	69,23	70	699,233
8	79	80	799
9	89	90	899
10	99,33,11	100	999,333,111
11	109,	200	1999
12	119,17	500	4999

If m has the values 40 or 400 instead of 4, it is necessary to work with two or three figures at a time instead of with one. Thus we find:

$$\frac{1}{399} = .\dot{0}0,\ 25,\ 06,\ 26,\ 56,\ 64,\ 16,\ 04,\ 0\dot{1}$$

$$\frac{1}{3999} = .\dot{0}00,\ 250,\ 062,\ 515,\ 628,\ 907,\ 226,\ \ldots\ldots$$

In this last case the process may be exhibited as below:

$$\begin{aligned} 1 &= 1 + 000 \times 4 \\ 1000 &= 0 + 250 \times 4 \\ 250 &= 2 + 062 \times 4 \\ 2062 &= 2 + 515 \times 4 \\ 2515 &= 3 + 628 \times 4 \\ 3628 &= 0 + 907 \times 4 \\ &\ldots\ldots \end{aligned}$$

Haverford College, Pa.
March, 1910.

CONCERNING THE INVARIANT POINTS OF COMMUTATIVE COLLINEATIONS.

By William Benjamin Fite.

1. In his *Geometrie der Lage** Reye states that a space collineation with just four invariant points is commutative only with the ∞^3 collineations that leave these points invariant.

The homogeneous linear substitution that represents such a collineation can have no two of its multipliers equal. Moreover if two collineations are commutative, the corresponding substitutions need not be commutative—one may transform the other into itself multiplied by a similarity substitution. If now in Reye's theorem we understand by commutative collineations those whose corresponding substitutions are commutative, the theorem is a special case of a much more general theorem. Without this restriction the theorem is not true, as I shall show.

If any substitution S transforms the substitution A into itself multiplied by a similarity substitution not identity, the sums of the multipliers of A and of S must be zero.† Reye's theorem applies then if the sum of the multipliers of A is not zero. The exception appears only in the case of collineations with "eingeschriebene Tetraederlage"‡, since a necessary and sufficient condition for such collineations is that the sum of the multipliers be zero.§

We suppose then that A has just four invariant points and that the sum of its multipliers is zero. It can be transformed to the normal form ¶

$$A \equiv \begin{pmatrix} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{pmatrix}, \quad (\omega_1 + \omega_2 + \omega_3 + \omega_4 = 0).$$

* Second part, 3rd edition, pp. 90, 91. Cf. *Schoenflies, Encyclopædie der Mathematischen Wissenschaften*, Band III, Heft 3, p. 468.

† Fite, *Transactions of the American Mathematical Society*, vol. 7 (1906), p. 66.

‡ R. Sturm, *Die Lehre von den geometrischen Vervandtschaften*, vol. III, p. 288.

§ After having proved this for any number of variables I found that the sufficiency of the condition for $n = 3$ had been given by Professor Morley in his lectures at Johns Hopkins University in 1902-3. The necessity of the condition is obvious.

¶ Cf. Weber, *Algebra*, vol. 2 (2nd edition), p. 175.

If S is a substitution such that $S^{-1}AS = AT$, where

$$T \equiv \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix},$$

ρ must be a fourth root of unity.

We shall treat the cases $\rho = -1$ and $\rho = i$ separately.*

2. $\rho = -1$. We can assume that A is of the form

$$A \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{pmatrix}, (\lambda \neq \pm 1).$$

In order that $S^{-1}AS = AT$, where

$$T \equiv \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

S must be of the form

$$S \equiv \begin{pmatrix} 0 & a_{12} & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{34} \\ 0 & 0 & a_{43} & 0 \end{pmatrix}.$$

The only invariant points of S are $(\sqrt{a_{12}}, \sqrt{a_{21}}, 0, 0)$, $(\sqrt{a_{12}}, -\sqrt{a_{21}}, 0, 0)$, $(0, 0, \sqrt{a_{34}}, \sqrt{a_{43}})$, and $(0, 0, \sqrt{a_{34}}, -\sqrt{a_{43}})$, provided that $a_{12}\, a_{21} \neq a_{34}\, a_{43}$. If this condition is not satisfied, S has invariant points in addition to these. If we denote the points $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$, by A_1, A_2, A_3, and A_4 respectively, it is clear that the first two of these invariant points of S lie on the line $A_1 A_2$ and are separated harmonically by A_1 and A_2; and that the last two lie on the line $A_3 A_4$ and are separated harmonically by A_3 and A_4.

* The case $\rho = -i$ is not essentially different from the case $\rho = i$.

Conversely, any point $(\rho_1, \rho_2, 0, 0)$, except A_1 and A_2, on $A_1 A_2$ and its harmonic conjugate $(\rho_1, -\rho_2, 0, 0)$ with respect to these two points, together with any point $(0, 0, \sigma_1, \sigma_2)$ except A_3 and A_4, on $A_3 A_4$ and its harmonic conjugate with respect to these two points are invariant points of collineations that are commutative with A. Such collineations are of the form

$$S \equiv \begin{pmatrix} 0 & \rho_1^2 & 0 & 0 \\ \rho_2^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 \\ 0 & 0 & \sigma_2^2 & 0 \end{pmatrix}.$$

If the σ's are multiplied by a common factor μ different from unity, the given invariant points are not affected, but we get a different collineation for S. Hence there is a single infinity of collineations commutative with A that have the given points for invariant points. Moreover the invariant points of A are not invariant under these collineations.

It is clear that any point on $A_1 A_2$ or $A_3 A_4$ is transformed by A into its harmonic conjugate with respect to A_1 and A_2 or A_3 and A_4 respectively. Hence: —

THEOREM. *Let A be any three-dimensional collineation with just four invariant points such that the points on two opposite edges of its invariant tetraedron are transformed by it into their harmonic conjugates with respect to the invariant points lying on these edges. Then there is a single infinity of collineations that are commutative with A and that have for invariant points any two points except the vertices on these two opposite edges and the harmonic conjugates of these points with respect to the corresponding vertices. Moreover these collineations do not leave invariant the invariant points of A.**

3. $\rho = i$. In this case $S^{-1}AS = AT$, where

$$T \equiv \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \end{pmatrix}.$$

* It is clear that the collineations A and S of this section can be so chosen that they will both put real points into real points and will both have real invariant tetraedra.

We can assume that

$$A \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -i \end{pmatrix},$$

and then, as may easily be verified,

$$S \equiv \begin{pmatrix} 0 & 0 & 0 & a_{14} \\ a_{21} & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 \\ 0 & 0 & a_{43} & 0 \end{pmatrix}.$$

The only invariant points of S are

$$\left(a_{14}, \frac{a_{21}a_{14}}{a}, \frac{a_{32}a_{21}a_{14}}{a^2}, a\right), \quad \left(a_{14}, \frac{ia_{21}a_{14}}{a}, -\frac{a_{32}a_{21}a_{14}}{a^2}, -ia\right),$$

$$\left(a_{14}, -\frac{a_{21}a_{14}}{a}, \frac{a_{32}a_{21}a_{14}}{a^2}, -a\right), \text{ and } \left(a_{14}, -\frac{ia_{21}a_{14}}{a}, -\frac{a_{32}a_{21}a_{14}}{a^2}, ia\right),$$

where $a = \sqrt[4]{a_{14}\, a_{43}\, a_{32}\, a_{21}}$

The first of these points is transformed into the others by A and its powers. Moreover we can determine S uniquely in such a way that the first of these points is any point not lying in a face of the invariant tetraedron of A, and the others are the points into which this point is transformed by A and its powers.*

This form of A is also a special case of the form considered in §2, and therefore there are collineations commutative with A with invariant points such as are described in that section.

These results can be formulated in the

THEOREM. *If A is a collineation with "eingeschriebene Tetraederlage," of period four, and with just four invariant points, which are taken as the ver-*

* It is clear that A is of period 4.

tices of the tetraedron of reference, and if $(\rho_1, \rho_2, \rho_3, \rho_4)$ *is any point not in a face of the tetraedron of reference, then*

$$S \equiv \begin{pmatrix} 0 & 0 & 0 & \rho_1 \\ \frac{\rho_2\rho_4}{\rho_1} & 0 & 0 & 0 \\ 0 & \frac{\rho_3\rho_4}{\rho_2} & 0 & 0 \\ 0 & 0 & \rho_3 & 0 \end{pmatrix}$$

is commutative with A *and has for its only invariant points the point* $(\rho_1, \rho_2, \rho_3, \rho_4)$ *and the points into which this is transformed by* A *and its powers.** *Moreover* S *is the only collineation that is commutative with* A *and that has these invariant points. The invariant points of* A *are not invariant under* S.

4. The general results of §§2 and 3 are applicable to collineations in n variables †$(n \geqq 2)$. Let d $(1 < d \leqq n)$ be any divisor of n and consider any simplex‡ with n vertices that is contained in a linear space of $n-1$ dimensions. We can select n/d linear bounding spaces of this simplex, each of $d-1$ dimensions, in such a way that no two of them have a common point. Then there are $\infty^{\frac{n}{d}-1}$ collineations A of period d that leave invariant the vertices of this simplex and no other points and that, together with their powers, transform any point in any one of these bounding spaces but not in any bounding space of fewer dimensions into d distinct points. If we select any one of these collineations A and a corresponding set of d such points in each of the n/d bounding spaces just referred to, there are $\infty^{\frac{n}{d}-1}$ collineations S that leave these points invariant and that are commutative with A. Moreover none of these collineations leave the invariant points of A invariant.

This result is a generalization of the theorem given by Stéphanos for§ $n = 2$.

---

* These points are obviously not co-planar.

† I am indebted to Professor Virgil Snyder for the suggestion to apply the general results of §§2 and 3 to collineations in n variables.

‡ Cf. Schoute, *Mehrdimensionale Geometrie*, vol. 1, p. 9.

§ *Mathematische Annalen*, vol. 22 (1883), p. 313.

5. The collineations commutative with two given commutative collineations. For one of the given collineations we shall take a collineation A in n variables that is of period n and that has just n invariant points. Such a collineation can be written in the form $A: x_i' = \omega^{i-1} x_i$ $(i = 1, 2, \cdots, n)$, where ω is a primitive nth root of unity. For the other given collineation we can take any collineation that is commutative with A. We shall consider first the collineation $S: x_i' = \rho^i x_{i-1}$ $(i = 1, 2, \cdots, n)$, where the subscripts are to be taken *modulo* n. Now a necessary and sufficient condition that a collineation be commutative with A is that it be of the form $x_i' = a_i x_{a+i}$. Hence S is commutative with A.

We consider now any collineation $S_1: x_i' = \sigma_i x_{d+i}$, where $0 \leqq d < n$, that is commutative with A, and determine under what conditions it is also commutative with S. Evidently $S_1 S$ is of the form $x_i' = \sigma_i \rho_{d+i} x_{d+i-1}$, while $S S_1$ is of the form $x_i' = \rho_i \sigma_{i-1} x_{d+i-1}$. Hence a necessary and sufficient condition that S and S_1 be commutative is that

$$\frac{\sigma_1 \rho_{d+1}}{\rho_1 \sigma_n} = \frac{\sigma_i \rho_{d+i}}{\rho_i \sigma_{i-1}}$$

for all values of i from 2 to n inclusive. For $i = 2$, we have*

$$\sigma_n = \frac{\rho_2 \rho_{d+1}}{\sigma_2 \rho_{d+2}}.$$

Moreover

$$\sigma_i = \frac{\sigma_2^{i-1} \rho_3 \rho_4 \cdots \rho_i \rho_{d+2}^{i-2}}{\rho_2^{i-2} \rho_{d+3} \rho_{d+4} \cdots \rho_{d+i}}.$$

If in this formula we put $i = n$ and equate the resulting value of σ_n with the one just found, we get just n distinct values for σ_2. But the value of σ_2 determines uniquely the values of all the other σ's. Hence there are exactly n collineations of the form S_1 that are commutative with S. But the n collineations $A^j S^{-d}$ $(j = 0, 1, \cdots, n-1)$ are all of this form and are all commutative with S (since A and S are commutative). We conclude therefore that *the only collineations that are commutative with both A and S are those contained in the abelian group of order n^2 generated by A and S.*

---

* For convenience we shall put ρ_1 and σ_1 each equal to unity. This obviously puts no restrictions on the collineations S and S_1.

If S is of the form $x_i' = \rho_i x_{d+i}$, where d is relatively prime to n, the result is the same as in the case just discussed. But if d is not relatively prime to n the situation is more complicated. If c is the greatest common divisor of n and d, and $n = rc$, then in order that a collineation S_1 of the form $x_i' = \sigma_i\, x_{e+i}$ be commutative with S it is necessary and sufficient that

$$\frac{\rho_1 \sigma_{d+1}}{\sigma_1 \rho_{e+1}} = \frac{\rho_{id+j}\sigma_{(i+1)d+j}}{\sigma_{id+j}\rho_{id+e+j}}$$

for all values of i from 0 to $r - 1$ inclusive and all values of j from 1 to c inclusive.† Putting $j = 1$, we get

$$\sigma_{d+1}^r = \frac{\rho_{e+1}^{r-1}\rho_{d+1} \cdots \rho_{id+1} \cdots \rho_{(r-1)d+1}}{\rho_{d+e+1} \cdots \rho_{id+e+1} \cdots \rho_{(r-1)d+e+1}} .$$

For $j = 2$ we get

$$\sigma_{d+1}^r = \frac{\rho_{e+1}^r \rho_2 \cdots \rho_{id+2} \cdots \rho_{(r-1)d+2}}{\rho_{e+2} \cdots \rho_{id+e+2} \cdots \rho_{(r-1)d+e+2}} .$$

If e is not a multiple of c, these two values of σ_{d+1}^r, when equated, give a relation connecting the coefficients of S.

Therefore in this case S is, in general, not commutative with any collineation of the form S_1.

But if e is a multiple of $c (e = sc)$, we have $\sigma_{d+1}^r = \rho_{sc+1}^r$. This gives r values for σ_{d+1}, and for each of these values there is a single set of values for σ_{id+1} $(i = 1, 2, \cdots, r - 1)$. In a similar way we get

$$\sigma_{d+j}^r = \frac{\sigma_j^r \rho_{sc+j}^r}{\rho_j^r} ,$$

and each of the resulting r values of σ_{d+j} (in terms of σ_j and the coefficients of S) determines a single set of values of σ_{id+j}. Moreover, inasmuch as

$$\frac{\rho_1 \sigma_{d+1}}{\sigma_1 \rho_{e+1}} = \frac{\rho_j \sigma_{d+j}}{\sigma_j \rho_{e+j}} ,$$

the value chosen for σ_{d+j} (in terms of σ_j) is uniquely fixed by the value chosen for σ_{d+1}. Hence *there are* ∞^{c-1} *collineations of the form* S_1 *that are commutative with both* A *and* S.

† If $d = r_1 c$ and x is a root of the congruence $r_1 x \equiv 1 \pmod{r}$, then $id + j \equiv i_1 d + c + j$. (mod n), where $i_1 \equiv i - x \pmod{r}$. Hence it is unnecessary to give to j values greater than c.

If c_1 is the greatest common divisor of e and n and $n = gc_1$, the order of S_1, if it is finite, is obviously a multiple of g. Now S_1^g is of the form $x'_{ie+j} = \beta_j x_{ie+j}$ $(i = 1, 2, \cdots, g;\ j = 1, 2, \cdots, c_1)$, where $\beta_j = \sigma_j \sigma_{e+j} \cdots \sigma_{le+j} \cdots \sigma_{(g-1)e+j}$. But $le + j \equiv i_1 d + j_1 \pmod{n}$ for a suitably chosen i_1, when j_1 is the least positive residue of j with respect to the modulus c. Hence $\sigma_{le+j} = \sigma_{j_1} R_l$ and $\beta_j = \sigma_{j_1}^g T_j$ where T_j and R_l are rational functions of the coefficients of S. For any S_1 whose order divides mg, $\sigma_{j_1}^{mg}$ $(j_1 = 1, 2, \cdots, c)$ is fixed. Moreover there are r choices for σ_{d+j_1} and for each of these choices there are mg values for each σ_{j_1}, except σ_1, which is equal to unity. Therefore *the number of collineations of the form S_1 that are commutative with both A and S and whose orders divide mg is $r(mg)^{c-1}$*.

Cornell University,
Ithaca, N. Y.
March, 1910.

A NEW CONSTRUCTION FOR CYCLOIDS

By H. Schapper

In the following lines is shown a new way for generating cycloids, and a simple method of constructing such curves. The advantage is looked for in the fact that, here, with the points of the curve are given simultaneously the respective tangents; and this is a constructional simplification. The kinematical aspect of the problem is also of interest.

Beginning with the simplest case,* consider a circle rolling uniformly over a straight line, and simultaneously a point P describing a simple harmonic motion (abbreviated SHM) along a diameter of the rolling circle. If to a complete revolution of the circle corresponds a complete period of the SHM, and if both motions begin at the same instant, and start from the same point, then the equations in rectangular coordinates of the path of P assume the form

$$x = r(a - \sin a \cos a),$$
$$y = r(1 - \cos^2 a),$$

where use is made of the relation that exists between SHM and uniform circular motion, so that P is found by dropping a perpendicular from the contact of the circle with the fixed line to the diameter in which the SHM takes place.

These equations may be written in the following way:

$$x = \frac{r}{2}(2a - \sin 2a),$$

$$y = \frac{r}{2}(1 - \cos 2a).$$

Putting $\frac{r}{2} = r'$, $2a = a'$, we finally get

$$x = r'(a' - \sin a'),$$
$$y = r'(1 - \cos a'),$$

* A proof for the construction in the case of the common cycloid was published by the author in *The American Mathematical Monthly*, February, 1909.

which is the common parameter form of the equations of the cycloid. We thus see that the cycloid may be defined as the path of a point having a *SHM* along the diameter of a uniformly rolling circle. The cycloid thus generated is the same as if described by a point on the rim of a rolling circle of radius $r/2$.

For the tangent we get

$$\frac{dy}{dx} = \frac{\sin 2a}{1 - \cos 2a} = \cot a = \tan\left(\frac{\pi}{2} - a\right),$$

which says that the diameter in which the *SHM* occurs is always tangent to the curve; the points of the curve as well as their corresponding tangents are thus found at the same time. This fact is also evident from kinematical considerations.

As the next case we consider the rolling of a circle of radius r on a fixed circle of radius R, and at the same time a *SHM* along a diameter of the moving circle, the conditions imposed being the same as in the case of the cycloid. Denoting by a the angle formed by the line of centers of the circles with its original direction, we get for a point P of the resulting path

$$x = (R + r)\cos a - r\cos\frac{R}{r}a\cos\frac{R + r}{r}a,$$

$$y = (R + r)\sin a - r\cos\frac{R}{r}a\sin\frac{R + r}{r}a.$$

After some reductions these equations simplify to the following:

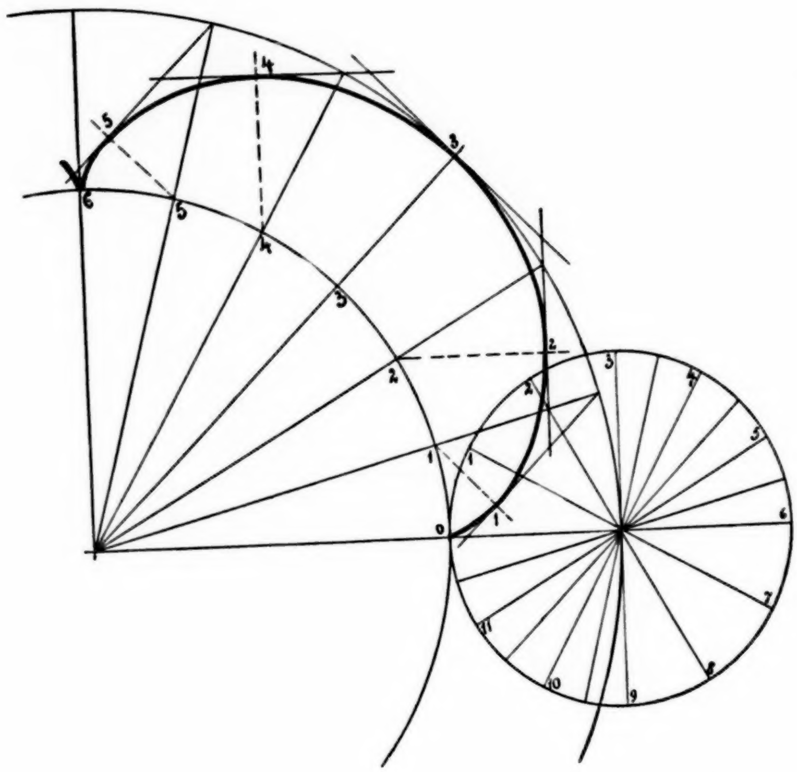
$$x = (R + r')\cos a - r'\cos\frac{R + r'}{r'}a,$$

$$y = (R + r')\sin a - r'\sin\frac{R + r'}{r'}a,$$

where

$$r' = \frac{r}{2}.$$

We thus see that the resulting path of a point describing a *SHM* under the stated conditions along the diameter of a circle of radius r rolling on another circle of radius R is an epicycloid. The epicycloid thus generated is



the same as if described by a point on the rim of a circle of radius $r/2$ rolling over one of radius R.

For the slope of the tangent we get

$$\frac{dy}{dx} = \tan \frac{R + r}{r} a = \tan \theta ,$$

where θ is the inclination of the diameter considered to Ox, and therefore the diameter in which the *SHM* takes place is tangent to the epicycloid for every

one of its points, so that we find simultaneously the points and the tangents to the curve.

In a similar manner we get for the case of a circle rolling on the inside of another circle, the *SHM* taking place along a diameter of the rolling circle under the same conditions as before,

$$x = (R - r)\cos a + r \cos \frac{R}{r} a \cos \frac{R - r}{r} a \,,$$

$$y = (R - r)\sin a - r \cos \frac{R}{r} a \sin \frac{R - r}{r} a,$$

and by reducing we get

$$x = (R - r')\cos a + r'\cos \frac{R - r'}{r'} a \,,$$

$$y = (R - r')\sin a - r'\sin \frac{R - r'}{r'} a,$$

where $r' = r/2$.

These equations express that the resulting path is an hypocycloid — the same as if described by a point on the rim of a circle of radius $r/2$ rolling over the inside of one of radius R.

For the tangent we get

$$\frac{dy}{dx} = - \tan \frac{R - r}{r} a = \tan \theta \,,$$

and therefore here, too, the diameter in which the *SHM* occurs is tangent to the hypocycloid for every one of its points.

The construction is similar to that in the case of the epicycloid.

It is also to be noticed that the points of the curve are here found *directly* as the intersection of the tangent and normal.

From the foregoing the conclusion seems to be justified that it is interesting and advantageous to define the cycloids as the resultant of a *SHM* combined with that of the rolling of a circle.

CARNEGIE TECHNICAL SCHOOLS,
PITTSBURGH, PA.

METRIC CLASSIFICATION OF CONICS AND QUADRICS BY MEANS OF RANK

By George Rutledge

It is well known that, under non-singular linear substitutions, the rank of the determinant of a homogeneous quadratic form, in n variables, is an absolute invariant. It follows directly from this that the rank of the determinant of a non-homogeneous quadratic form, in $n-1$ variables, and the rank of the determinant of its homogeneous quadratic part, are absolute invariants under substitutions of the type

$$x_i = l_{i,1}\, y_1 + l_{i,2} y_2 + \cdots + l_{i,\, n-1}\, y_{n-1} + l_{i,\, n} \qquad (i = 1, 2, \cdots, n-1),$$

where the determinant of the homogeneous part of the substitution is non-singular.

Hence, in particular, the rank of the determinant of the equation of a conic or quadric, and the rank of the determinant of its homogeneous quadratic part are absolute invariants under all rigid displacements of the conic or quadric. This fact may be made use of to determine the type of a conic or quadric by a mere inspection of its equation, without any transformation.

It is shown in the text-books,* that the equation of a quadric may always be reduced, by rigid displacements, to the form,

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 + 2kx_3 + c = 0,$$

where either k or λ_3 is zero, and either k or c is zero.

* For instance, Fort-Schlömilch, *Lehrbuch der analytischen Geometrie*, vol. 2, p. 207.
For a purely analytic treatment of this reduction, cf. Bromwich. *Quadratic Forms and their Classification by Means of Invariant Factors* (Cambridge, England, 1906), p. 72.

There are, then, three types of the reduced equation, as follows:

First type: Reduced equation, $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 + c = 0, \quad c \neq 0$

I_1	$\lambda_1 \cdot \lambda_2 \cdot \lambda_3 \neq 0$	Central Quadric
I_2	$\lambda_1 \cdot \lambda_2 \neq 0,\ \lambda_3 = 0$	Elliptic or Hyperbolic Cylinder
I_3	$\lambda_1 \neq 0,\ \lambda_2 = 0,\ \lambda_3 = 0$	Two Parallel Planes

Second type: Reduced equation, $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 = 0$

II_1	$\lambda_1 \cdot \lambda_2 \cdot \lambda_3 \neq 0$	Cone
II_2	$\lambda_1 \cdot \lambda_2 \neq 0,\ \lambda_3 = 0$	Two Intersecting Planes
II_3	$\lambda_1 \neq 0,\ \lambda_2 = 0,\ \lambda_3 = 0$	Two Coincident Planes

Third type: Reduced equation, $\lambda_1 x_1^2 + \lambda_2 x_2^2 + 2kx_3 = 0, \ k \neq 0$

III_2	$\lambda_1 \cdot \lambda_2 \neq 0,\ \lambda_3 = 0$	Paraboloid
III_3	$\lambda_1 \neq 0,\ \lambda_2 = 0,\ \lambda_3 = 0$	Parabolic Cylinder

If, in each case, we determine the rank of the determinant

$$\begin{vmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & k \\ 0 & 0 & k & c \end{vmatrix}$$

this rank will be the same as the rank of the corresponding determinant of the original unreduced equation, and similarly for the rank of the principal first minor in the upper left-hand corner.

The minors which do not vanish identically are:

$$\begin{vmatrix} \lambda_2 & 0 & 0 \\ 0 & \lambda_3 & k \\ 0 & k & c \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_3 & k \\ 0 & k & c \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & c \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & k \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{vmatrix}$$

$$\begin{vmatrix} \lambda_2 & 0 \\ 0 & \lambda_3 \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_3 \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 \\ 0 & c \end{vmatrix} \quad \begin{vmatrix} \lambda_1 & 0 \\ 0 & k \end{vmatrix} \quad \begin{vmatrix} \lambda_2 & 0 \\ 0 & c \end{vmatrix} \quad \begin{vmatrix} \lambda_2 & 0 \\ 0 & k \end{vmatrix} \quad \begin{vmatrix} \lambda_3 & k \\ k & c \end{vmatrix}$$

An examination of the above determinant and minors yields the following result, where $R =$ the rank of the determinant of the *original* unreduced equation, and $R_1 =$ the rank of its principal first minor in the upper left-hand corner:

R	R_1	TYPE
4	3	I_1 Central Quadric
4	2	III_2 Paraboloid
3	3	II_1 Cone
3	2	I_2 Elliptic or Hyperbolic Cyl.
3	1	III_3 Parabolic Cylinder
2	2	II_2 Two Intersecting Planes
2	1	I_3 Two Parallel Planes
1	1	II_3 Two Coincident Planes
1	0	Equation linear

The necessary changes in the case of conics will be recognized without difficulty. We obtain:

3	2	I_1 Ellipse or Hyperbola
3	1	III_2 Parabola
2	2	II_1 Two Intersecting Lines
2	1	I_2 Two Parallel Lines
1	1	II_2 Two Coincident Lines
1	0	Equation linear

As an example of the foregoing method, consider the quadric*

* Discussed by the ordinary method in C. Smith, *Solid Geometry*, p. 63.

$$4x^2 + y^2 + 4z^2 - 4yz + 8zx - 4xy + 2x - 4y + 5z + 1 = 0.$$

The determinant to be considered is

$$\begin{vmatrix} 4 & -2 & 4 & 1 \\ -2 & 1 & -2 & -2 \\ 4 & -2 & 4 & \frac{5}{2} \\ 1 & -2 & \frac{5}{2} & 1 \end{vmatrix}$$

Adding twice the second column to the first and to the third, which does not change the rank of either of the determinants with which we are concerned,† we have,

$$\begin{vmatrix} 0 & -2 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & -2 & 0 & \frac{5}{2} \\ -3 & -2 & -\frac{3}{2} & 1 \end{vmatrix}$$

whence it is evident that the determinant of the original equation is of rank 3, and that its principal first minor in the upper left-hand corner is of rank 1.

The quadric represented is consequently a parabolic cylinder.

THE UNIVERSITY OF ILLINOIS,
APRIL, 1910.

† Böcher, *Introduction to Higher Algebra*, p. 55.

A METHOD OF SOLVING LINEAR DIFFERENTIAL EQUATIONS*

By P. A. Lambert

The object of this paper is to present a new method of solving ordinary linear diffential equations, which may frequently be applied with advantage when the coefficients of the equation are polynomials in the independent variable.

Let the given differential equation be

$$(1)\qquad f\left(x,\, y,\, \frac{dy}{dx},\, \frac{d^2y}{dx^2},\, \cdots \frac{d^ny}{dx^n}\right) = 0.$$

The method of solution proposed consists of the following steps:

(a) Break up the function f into two parts, one of which, f_1, equated to zero gives a differential equation which may be readily solved, and introduce a parameter t as a factor of the second part so that the given equation, $f_1 + f_2 = 0$, is replaced by

$$(2)\qquad f_1 + tf_2 = 0.$$

(b) Assume that the series

$$(3)\qquad y = A + Bt + Ct^2 + Dt^3 \cdots,$$

where A, B, C, D, $\cdots$ are undetermined functions of x, satisfies (2). Substitute the expression (3) for y in (2) and determine these functions by solving the differential equations formed by equating to zero the coefficients of successive powers of t in this identity.

(c) Substitute the values of A, B, C, D, $\cdots$ in (3), and replace t by unity, and see whether

* Read before the American Mathematical Society, October 30, 1909.

$$(4) \qquad y = A + B + C + D \cdots$$

satisfies equation (1).

It is to be noted that A is a solution of the equation $f_1 = 0$, and may consequently be frequently determined by a suitable choice of f_1 to contain n arbitrary constants so leading to the general solution of (1). If this is possible it is necessary only to obtain particular solutions of the equations for $B, C, D, \cdots$.

Since the process above described is purely formal it is evidently necessary to see whether or not the series (4) actually satisfies (1) if that series contains an infinite number of terms, or if any of the functions $A, B, C, \cdots$ is given by an infinite series.

The most advantageous method of breaking up the given equation into two parts must be determined by trial. However, if no term is separated into two parts, the number of possible methods of choosing f_1 and f_2 is never greater than 2^n, and if n is not large the best method may be selected without much difficulty.

In the process of solving the differential equations which determine $A, B, C, D, \cdots$, independent arbitrary constants are introduced until the number of arbitrary constants equals the order of the given differential equation.

If the independent arbitrary constants are $C_1, C_2, C_3, \cdots C_n$ the terms of the series (4) may be grouped so that (4) takes the form

$$(5) \qquad y = C_1 y_1 + C_2 y_2 + C_3 y_3 + \cdots + C_n y_n + Y,$$

where $y_1, y_2, y_3, \cdots y_n$ are independent solutions of the corresponding homogeneous differential equation, and Y is a particular integral of (1).

If another manner of breaking up the given differential equation makes series (4) a solution of (1), there may be determined a different set of independent integrals and a different particular integral.

If the general solution (5) of the differential equation contains infinite series, the limits of the regions of convergency must be determined by the usual methods.

This method of solving differential equations is the result of an attempt to extend to differential equations the method employed by the author in the papers "On the solution of algebraic equations in infinite series." †

† *Bulletin American Mathematical Society*, ser. 2, vol. 14, 1908, pp. 467-477. *Proceedings American Philosophical Society*, vol. 47, 1908, pp. 111-134.

The incentive to make this attempt came from the following statement in an extract of a letter from Cauchy to Coriolis of January 29, 1837, published in the *Comptes Rendus* of the Paris Academy.

"Ainsi étendus, ces mèthodes s'appliquent avec un succès remarquable à presque tous les grands problèmes d'analyse, à la résolution générale des équations, à l'intégration des équations différentielles, à la mécanique céleste, etc."

Cauchy describes the method applied to algebraic equations as follows:

"Pour résoudre une équation partagez son premier membre en deux polynomes d'une manière quelconque, et supposez l'un de ces polynomes multiplié par un paramètre que vous reduisez plus tard à l'unité."

By this method the solutions of Bessel's equations, of Legendre's equation, and of the differential equation of the hypergeometric series may be advantageously determined.

The method will be examplified by applying it to two differential equations.

Example 1. Solve

$$\frac{d^2y}{dx^2} + ax^2y = 1 + x.$$

Writing this equation in the form

$$\frac{d^2y}{dx^2} - (1 + x) + ax^2yt = 0,$$

and assuming that

$$y = A + Bt + Ct^2 + Dt^3 + \cdots$$

there results

$$\left.\begin{array}{r} \dfrac{d^2A}{dx^2} \\ -(1+x) \end{array}\right| + \left.\begin{array}{c} \dfrac{d^2B}{dx^2} \\ + ax^2A \end{array}\right| t + \left.\begin{array}{c} \dfrac{d^2C}{dx^2} \\ + ax^2B \end{array}\right| t^2 + \left.\begin{array}{c} \dfrac{d^2D}{dx^2} \\ + ax^2C \end{array}\right| t^3 + \cdots \equiv 0.$$

Equating to zero the coefficients of the successive powers of t in this identity, we have

$$\frac{d^2A}{dx^2} = 1 + x,$$

$$\frac{d^2B}{dx^2} + ax^2A = 0,$$

$$\frac{d^2C}{dx^2} + ax^2B = 0,$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

From this series of differential equations

$$A = K_1 + K_2x + \frac{x^2}{1\cdot 2} + \frac{x^3}{2\cdot 3},$$

$$B = -K_1\frac{ax^4}{3\cdot 4} - K_2\frac{ax^5}{4\cdot 5} - \frac{ax^6}{1\cdot 2\cdot 5\cdot 6} - \frac{ax^7}{2\cdot 3\cdot 6\cdot 7},$$

$$C = K_1\frac{a^2x^8}{3\cdot 4\cdot 7\cdot 8} + K_2\frac{a^2x^9}{4\cdot 5\cdot 8\cdot 9} + \frac{a^2x^{10}}{1\cdot 2\cdot 5\cdot 6\cdot 9\cdot 10} + \frac{a^2x^{11}}{2\cdot 3\cdot 6\cdot 7\cdot 10\cdot 11},$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

The law of formation of the successive coefficients A, B, C, D, $\cdots$ is evident, and the value of y, when t is made unity, becomes

$$\begin{aligned} y = K_1 &\left(1 - \frac{ax^4}{3\cdot 4} + \frac{a^2x^8}{3\cdot 4\cdot 7\cdot 8} - \frac{a^3x^{12}}{3\cdot 4\cdot 7\cdot 8\cdot 11\cdot 12} + \cdots\right) \\ + K_2x &\left(1 - \frac{ax^4}{4\cdot 5} + \frac{a^2x^8}{4\cdot 5\cdot 8\cdot 9} - \frac{a^3x^{12}}{4\cdot 5\cdot 8\cdot 9\cdot 12\cdot 13} + \cdots\right) \\ + \frac{x^2}{2} &\left(1 - \frac{ax^4}{5\cdot 6} + \frac{a^2x^8}{5\cdot 6\cdot 9\cdot 10} - \cdots\right) \\ + \frac{x^3}{6} &\left(1 - \frac{ax^4}{6\cdot 7} + \frac{a^2x^8}{6\cdot 7\cdot 10\cdot 11} - \cdots\right). \end{aligned}$$

This value of y is composed of four infinite series, each convergent for all values of x, and is the complete solution of the given differential equation.

Example 2. Solve

$$x^2\frac{d^2y}{dx^2} + (x + 2x^2)\frac{dy}{dx} - 4y = 0\cdot$$

Writing this equation in the form

$$x^2 \frac{d^2y}{dx^2} + x \frac{dy}{dx} - 4y + 2x^2t \frac{dy}{dx} = 0,$$

and assuming that

$$y = A + Bt + Ct^2 + Dt^3 + Et^4 + \cdots ,$$

there results

$$\left. \begin{array}{r} x^2 \dfrac{d^2A}{dx^2} \\ + x \dfrac{dA}{dx} \\ - 4A \end{array} \right| + \left. \begin{array}{r} x^2 \dfrac{d^2B}{dx^2} \\ + x \dfrac{dB}{dx} \\ - 4B \\ + 2x^2 \dfrac{dA}{dx} \end{array} \right| t + \left. \begin{array}{r} x^2 \dfrac{d^2C}{dx^2} \\ + x \dfrac{dC}{dx} \\ - 4C \\ + 2x^2 \dfrac{dB}{dx} \end{array} \right| t^2 + \cdots \equiv 0.$$

Equating to zero the coefficients of the successive powers of t in this identity, we have

$$x^2 \frac{d^2A}{dx^2} + x \frac{dA}{dx} - 4A = 0,$$

$$x^2 \frac{d^2B}{dx^2} + x \frac{dB}{dx} - 4B + 2x \frac{dA}{dx} = 0,$$

$$x^2 \frac{d^2C}{dx^2} + x \frac{dC}{dx} - 4C + 2x^2 \frac{dB}{dx} = 0,$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$

The solutions of the first equation for A are $A = K_1x^2$ and $A = K_2x^{-2}$.
(a) Substituting $A = K_1x^2$ in the next equation,

$$x^2 \frac{d^2B}{dx^2} + x \frac{dB}{dx} - 4B + 4K_1 x^3 = 0.$$

A particular solution of this differential equation must be determined. The particular solution may be found by multiplying this equation by x and by two successive direct integrations.

The particular solution may also be found by the method of this paper as follows: Writing the equation in the form

$$x^2 \frac{d^2B}{dx^2} + x\frac{dB}{dx} + 4K_1x^3 - 4Bt = 0$$

and assuming that

$$B = B_1 + B_2t + B_3t^2 + B_4t^3 + \cdots,$$

there results

$$\begin{array}{l|l|l|l}
x^2 \dfrac{d^2B_1}{dx^2} & + \; x^2 \dfrac{d^2B_2}{dx^2} & t + x^2 \dfrac{d^2B_3}{dx^2} & t^2 + \cdots \equiv 0. \\
+ \; x\dfrac{dB_1}{dx} & + \; x\dfrac{dB_2}{dx} & + \; x\dfrac{dB_3}{dx} & \\
+ \; 4K_1x^3 & - \quad 4B_1 & - \quad 4B_2 &
\end{array}$$

Equating to zero the coefficients of the successive powers of t in this identity, we have

$$x^2 \frac{d^2B_1}{dx^2} + x\frac{dB_1}{dx} + 4K_1x^3 = 0,$$

$$x^2 \frac{d^2B_2}{dx^2} + x\frac{dB_2}{dx} - 4B_1 = 0,$$

$$x^2 \frac{d^2B_3}{dx^2} + x\frac{dB_3}{dx} - 4B_2 = 0,$$

$$\cdots \quad \cdots \quad \cdots \quad \cdots$$

From this series of differential equations, we find the particular solutions

$$B_1 = -\frac{4}{9}K_1x^3,\; B_2 = -\left(\frac{4}{9}\right)^2 K_1x^3,\; B_3 = -\left(\frac{4}{9}\right)^3 x^3,\; \cdots .$$

Hence

$$B = -\frac{4}{9}K_1x^3 - \left(\frac{4}{9}\right)^2 K_1x^3 - \left(\frac{4}{9}\right)^3 K_1x^3 - \cdots = -\frac{4}{5}K_1x^3.$$

Substituting this value of B,

$$x^2 \frac{d^2C}{dx^2} + x\frac{dC}{dx} - 4C - \frac{2\cdot3\cdot4}{5}K_1x^4 = 0.$$

A particular solution of this equation is

$$C = \frac{3 \cdot 4}{5 \cdot 6} K_1 x^4.$$

In like manner

$$D = -\frac{4 \cdot 8}{5 \cdot 6 \cdot 7} K_1 x^5, \quad E = \frac{5 \cdot 16}{5 \cdot 6 \cdot 7 \cdot 8} K_1 x^6, \cdots$$

$$N = (-1)^n \frac{(n+1)\, 2^n}{5 \cdot 6 \cdot 7 \cdot 8 \cdots (n+4)} x^{n+2}.$$

Substituting the values of A, B, C, D, $\cdots$, and making t unity,

$$y = K_1 x^2 \Big[1 - \frac{4}{5} x + \frac{3 \cdot 4}{5 \cdot 6} x^2 + \frac{4 \cdot 8}{5 \cdot 6 \cdot 7} x^3 + \cdots$$

$$+ (-1)^n \frac{(n+1)\, 2^n}{5 \cdot 6 \cdot 7 \cdot 8 \cdots (n+4)} x^n \cdots,$$

which is convergent for all values of x and a solution of the given differential equation.

(*b*) Substituting $A = K_2\, x^{-2}$,

we have

$$x^2 \frac{d^2 B}{dx^2} + x \frac{dB}{dx} - 4B - 4K_2 x^{-1} = 0.$$

A particular solution of this equation, found by the same methods, is

$$B = -\frac{4}{3} K_2 x^{-1}.$$

Substituting this value of B

$$x^2 \frac{d^2 C}{dx^2} + x \frac{dC}{dx} - 4C + \frac{8}{3} K_2 = 0.$$

A particular solution of this equation is

$$C = \frac{2}{3} K_2.$$

It is evident that $D = 0$, $E = 0$ $\cdots$ are particular solutions of the remaining differential equations of the series.

Hence

$$y = K_2 \left(x^{-2} - \frac{4}{3} x^{-1} + \frac{2}{3} \right)$$

is a solution of the given differential equation.

The general solution of the given differential equation is the sum of these two independent integrals.

LEHIGH UNIVERSITY,
SOUTH BETHLEHEM, PA.
FEBRUARY, 1910.

INDEX

OCTOBER 1909

ANNALS OF MATHEMATICS

(FOUNDED BY ORMOND STONE)

PUBLISHED UNDER THE AUSPICES OF HARVARD UNIVERSITY

SECOND SERIES Vol. 11 No. 1

FOR SALE BY
THE PUBLICATION OFFICE OF HARVARD UNIVERSITY
2 University Hall, Cambridge, Mass., U. S. A.

London: LONGMANS, GREEN & CO.
39 Paternoster Row

Leipzig: OTTO HARRASSOWITZ
Querstrasse 14

The following pamphlets, most of which have been reprinted from the ANNALS OF MATHEMATICS, second series, will be sent post-paid on receipt of price by the PUBLICATION OFFICE OF HARVARD UNIVERSITY, 2 University Hall, Cambridge, Mass.

MAXIME BÔCHER, Regular Points of Linear Differential Equations of the Second Order. 23 pp. 1896 . . . . $0.50

W. F. OSGOOD, Introduction to Infinite Series. 71 pp. Second Edition, 1902. . . .75

JAMES PIERPONT, Galois' Theory of Algebraic Equations. 67 pp. 1900. *From vols.* 1–2. .75

MAXIME BÔCHER, The Theory of Linear Dependence. 16 pp. 1901. *From vol.* 2. . .35

W. F. OSGOOD, Sufficient Conditions in the Calculus of Variations. 25 pp. 1901. *From vol.* 2. 50

F. S. WOODS, Space of Constant Curvature. 42 pp. 1902. *From vol.* 3. 50

W. F. OSGOOD, The Integral as the Limit of a Sum, and a Theorem of Duhamel's 18 pp. 1903. *From vol.* 4. 35

E. V. HUNTINGTON, The Continuum as a Type of Order: An Exposition of the Modern Theory. With an Appendix on the Transfinite Numbers. 63 pp. 1905. *From vols.* 6–7. 50

MAXIME BÔCHER, Introduction to the Theory of Fourier's Series 72 pp. 1906. *From vol.* 7. 75

E. V. HUNTINGTON, The Fundamental Laws of Addition and Multiplication in Elementary Algebra. 44 pp. 1906. *From vol.* 8 50

CONTENTS

ANNALS OF MATHEMATICS

Published in October, January, April, and July, under the auspices of Harvard University, Cambridge, Mass., U. S. A.

Cambridge: Address *The Annals of Mathematics*, 2 University Hall, Cambridge, Mass., U. S. A. Subscription price, $2 a volume (four numbers) in advance. Single numbers, 75c. All drafts and money orders should be made payable to Harvard University.

London: Longmans, Green & Co., 39 Paternoster Row. Price, 2 shillings a number.

Leipzig: Otto Harrassowitz, Querstrasse 14. Price, 2 marks a number.

Printed by the Salem Press Co., Salem, Mass., U. S. A.

10

m. o.